Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung
Certificate
Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

97120867.3

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Europäisches Patentamt | European Patent Office | Office européen des brevets

Anmeldung Nr:
Application no.: 97120867.3
Demande no:

Anmeldetag:
Date of filing: 27.11.97
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V.

14195 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Improved method for the identification and characterization of interacting molecules

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

Max-Planck-Gesellschaft zur Förderung
der Wissenschaften e. V.
Our Ref.: B 2021 EP

IMPROVED METHOD FOR THE IDENTIFICATION AND CHARACTERIZATION OF INTERACTING MOLECULES

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to detect positive clones from the rather large numbers of false positive clones isolated by conventional two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules from false positive clones. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules.

Protein-protein interactions are essential for nearly all biological processes like replication, transcription, secretion, signal transduction and metabolism. Classical methods for identifying such interactions like co-immunoprecipitation or cross-linking are not available for all proteins or may not be sufficiently sensitive. Said methods further have the disadvantage that only by a great deal of energy, potentially interacting partners and corresponding nucleic acid fragments or sequences may be identified. Usually, this is effected by protein sequencing or production of antibodies, followed by the screening of an expression-library.

An important development for the convenient identification of protein-protein interactions was the yeast two-hybrid (2H) system presented by Fields and Song (1989).This genetic procedure not only allows the rapid demonstration of in vivo interactions, but also the simple isolation of corresponding nucleic acid sequences encoding for the interacting partners. The yeast two-hybrid system makes use of the features of a wide variety of eukaryotic transcription factors which carry two separable functional domains: one DNA binding domain as well as a second domain which activates the RNA-polymerase complex (activation domain). In the classical 2H system a so-called "bait" protein comprising of a DNA binding domain (GAL4bd or lex A) and a protein of interest „X" are

expressed as a fusion protein in yeast. The same yeast cell also simultaneously expresses a so called "fish" protein comprising of an activation domain (GAL4ad or VP16) and a protein „Y". Upon the interaction of a bait protein with a fish protein, the DNA binding and activation domains of the fusion proteins are brought into close proximity and the resulting protein complex triggers the expression of the reporter genes, for example, HIS3 or lacZ. Said expression can be easily monitored by cultivation of the yeast cells on selective medium without histidine as well as upon the activation of the lacZ gene. The genetic sequence encoding, for example, an unknown fish protein, may easily be identified by isolating the corresponding plasmid and subsequent sequence analysis. Meanwhile, a number of variants of the 2H system have been developed. The most important of those are the "one hybrid" system for the identification of promoter binding proteins and the "tri-hybrid" system for the identification of RNA-protein-interactions (Li and Herskowitz, 1993; SenGupta et al., 1996; Putz et al., 1996).

The classical 2H system for the identification of protein-protein-interaction, has, until today, only been carried out on a laboratory scale. The various steps of this system need to be conducted serially. They are, therefore, quite time consuming. As a consequence, the 2H system has so far proven unsuitable for the analysis of eukaryotic library vs library screens to investigate protein-protein networks. Although recent developments have taken into account these disadvantages (Bartel et al.,1996), a successful large scale search of interacting proteins, for example on the basis of a eukaryotic library vs. library screen, has not been reported. More importantly, also all of the so far developed 2H systems suffer from the serious drawback that many false-positive clones not representing any interactions between binding partners are isolated. This is particularly inconvenient in cases where large numbers of clones are to be analyzed because in the case of a eukaryotic library vs library screen it is typical that several hundreds of thousands of clones have to be analyzed for the investigation of protein-protein networks.

The technical problem underlying the present invention was therefore to overcome these prior art difficulties and to furnish a system that reliably produces clones that express interacting molecules. This system should, moreover, be suitable for large-scale library vs library screens using a parallel, high-throughput or automated approach.

The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a method for the identification of at least one member of a pair or complex of interacting molecules, comprising:

(a) providing host cells containing at least two genetic elements with different selectable and counterselectable markers, said genetic elements each comprising genetic information specifying one of said members, said host cells further carrying a readout system that is activated upon the interaction of said molecules;

(b) allowing at least one interaction, if any, to occur;

(c) selecting for said interaction by transferring progeny of said host cells to

(ca) at least two different selective media, wherein each of said selective media allows growth of said host cells only in the absence of at least one of said counterselectable markers and in the presence of a selectable marker; and

(cb) a further selective medium that allows identification of said host cells only on the activation of said readout system ;

(d) identifying host cells containing interacting molecules that

(da) do not activate said readout system on any of said selective media specified in (ca); and

(db) activate the readout system on said selective medium specified in (cb); and

(e) identifying at least one member of said pair or complex of interacting molecules.

Preferably, said interaction is a specific interaction.

The terms "identification" and "identifying", as used in accordance with the present invention, relate to the ability of the person skilled in the art to detect positive clones that express interacting molecules from false positive clones due to the activation of the readout system on the selective media and optionally additionally to characterize at least one of said interacting molecules by one or a set of unambiguous features. Preferably, said molecules are characterized by the DNA sequence encoding them, upon nucleic acid hybridization or isolation and sequencing of the respective DNA molecules. Alternatively and less preferred, said molecules may be characterized by different features such as molecular weight, isoelectric point and, in the case of proteins, the N-terminal amino acid sequence etc. Methods for determining such parameters are well known in the art.

Preferably, said members specified by said genetic elements are connected to a further entity that will upon the interaction activate or contribute to the activation of said read out system. It is further preferred that said entity is conserved for each type of genetic element and that different types of genetic elements comprise different entities. It is additionally preferred that said member of said pair or complex of interacting molecules forms, when transcribed as RNA from said genetic element, an RNA transcript fused with RNA specifying said entity. Most preferably, said fused RNA transcript is translated to form a fusion protein comprising said member fused to said entity. As will be elaborated further herein below, said entity may be in one type of genetic element a DNA sequence encoding a DNA-binding domain and in a different type of genetic element a transactivating protein domain. Preferably, said genetic elements are vectors such as plasmids. The at least two genetic elements comprised in said host cell are preferentially vectors from a library such as a cDNA or genomic library. Thus, the method of the invention allows the screening of a variety of host cells wherein the vector portion of said genetic elements is preferably the same for each type of genetic element whereas the potentially interacting molecules are representatives of a library and, thus, as a rule and in case that the library has not been amplified, may differ in each host

cell. In this connection the term "type of genetic element" refers to an element characterized by comprising the same entity, selectable and counterselectable markers.

Preferably, the "interaction" of said molecules is specific and characterized by a high binding constant. However, the term "interaction" may also refer to a binding between molecules with a lower binding constant which, however, must be sufficient to activate the readout system. The interaction that is detectable by the method of the invention preferably leads to the formation of a functional entity having a biological, physical or chemical activity which was not present in said host cell before said interaction occurred.

Said interaction may preferably lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety that drives the activation of said readout system. For example, said moiety may be a promoter. Alternatively, said interaction may lead to a detectable fluorescence resonance energy transfer obtained by the interaction of fusion proteins containing, for example, the GFP type a and GFP type b fluorescent proteins (Cubbitt et al., 1995).

In a further embodiment, said interaction may lead to a detectable modification of a substrate by an enzyme such as a color reaction obtained by the cleavage of a propeptide by an enzyme. In all these embodiments of the invention, it is understood that the interacting molecules are preferably directly fused to the molecules driving the readout system.

The term "growth" on selective media "in the absence of at least one of said counter-selectable markers" refers to the fact that a population of host cells containing at least a pair of genetic elements is placed on said selective media but only those progeny of the host cells in the overall population that have lost the relevant genetic element are able to grow. For example, when a yeast strain which is resistant to the drug canavanine (can^r) and which also contains a

plasmid carrying the wild-type CAN1 gene (Hoffmann, 1985) is placed on a selective medium containing canavanine, only those progeny of the yeast strain that have lost the plasmid carrying the CAN1 gene are able to grow, because this gene confers sensitivity to canavanine in yeast cells.

With reference to step (ca), it should be noted that each of the at least two selective media would comprise at least one counterselectable compound such as cycloheximide wherein the counterselectable compound would be different in the different selective media; they would further typically lack a compound complementing for an auxotrophic marker or comprise an antibiotic. The compound or antibiotic may be the same for the various selective media. Preferably, at least one is different.

The method of the present invention provides a highly effective tool for selecting against false positive clones that have proven to dramatically reduce the overall usefulness of the two-hybrid system. For example, by inclusion of a marker counterselecting for the absence of a genetic element that specifies one of a pair of the potentially interacting partners, clones that will grow and therefore only carry the second genetic element specifying the second partner can now be tested for the activation of the readout system. If the clone containing only the fusion protein encoded by the second genetic element activates the readout system in the absence of the other genetic element, then it will be classified as a false positive. By counterselecting for the absence of the second genetic element, the same test is applied to the first genetic element. Thus, only clones that activate the readout system in the presence of both or all genetic elements, but do not activate the read out system when either of the genetic elements is lost are classified as positives.

The advantages associated with the method of the invention have a significant impact in particular on the number of clones that express potentially interacting partners that can conveniently be analyzed. For example, even work on the laboratory scale will be more effective since positive clones that express interacting partners can be easily and unambiguously discriminated from false positive clones without the generation of additional strains. In contrast, to detect

false positive clones using the state of the art yeast two-hybrid system, plasmids that encode fish proteins usually need to be isolated and retransformed into yeast cells harboring plasmids that encode unrelated bait proteins. Further, the enormous number of false positive clones that would be isolated when using the classical two-hybrid system on a large scale, yet are discriminated by the method of this invention no longer precludes an effective high through-put analysis of clones. In the long run, it is expected that the method of the present invention is especially advantageous for a high throughput analysis of a large number of yeast clones containing interacting molecules since many specific interactions and the individual members of these interactions can be identified in a parallel and automated approach.

Some investigators have noted the problem of identifying false positive clones when applying the yeast two-hybrid system in the past. Bartel et al. (1996) described a method for the elimination of false positives by replica plating clones that express one fusion protein from SD-leu and SD-trp plates, to SD-his plates. Clones that showed growth on the SD-his plates where identified as false positives and were subsequently not used for interaction mating. The disadvantage of this method is that the procedure is labor intensive because yeast strains expressing the fish proteins, the bait proteins and the potentially interacting fish and bait proteins all must be generated and analyzed. The use of the counterselectable system described in this invention has the advantage that only one strain which expresses the potentially interacting fusion proteins is generated and must be analyzed.

A schematic overview of one embodiment of the method of the invention is provided in Figure 1. For the parallel analysis of a network of protein-protein interactions with the method of the invention, a library of plasmid constructs that express DNA binding domain and activation domain fusion proteins is provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into the improved binding domain and activating domain plasmids of the invention containing different selectable and counterselectable markers. Both libraries are combined within yeast cells by

transformation or interaction mating, and yeast strains that express potentially interacting proteins are selected on selective medium lacking histidine. The selective markers TRP1 and LEU2 maintain the plasmids in yeast strains grown on selective media, whereas CAN1 and CYH2 specify the counter-selectable markers that select for the loss of each plasmid. HIS3 and lacZ represent selectable markers integrated into the yeast genome, which are expressed on activation by interacting fusion proteins.

The readout system is, in the present case, both growth on medium lacking histidine and enzymatic activity of ß-galactosidase which can be subsequently screened. It is to be understood, however, that the readout system may rely on only one marker such as HIS3. Yet, the combination of two components that constitute the readout system in many cases allows a more ready interpretation of results, in particular if one of the components, when activated, effects a change in color. A colony picking robot is used to pick the resulting yeast colonies into individual wells of 384-well microtiter plates containing selective medium lacking histidine, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library contained in microtiter plates can be optionally replicated and stored. The resulting interaction library is investigated to detect positive clones that express interacting proteins and discriminate them from false positive clones using the method of the invention. Using a spotting robot, cells are transferred to replica membranes which are subsequently placed onto one each of the selective media SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX. After incubation on the selective plates, the clones grown on the membranes are subjected to a ß-Gal assay and a digital image from each membrane is obtained with a CCD camera which is then stored on computer. Using digital image processing and analysis (Lehrach et al. 1997) clones that express interacting fusion proteins can be identified by considering the pattern of ß-Gal activity from clones grown on the various selective media. The individual members comprising interactions can then be identified by one or more techniques, including PCR, sequencing, hybridization, oligofingerprinting or antibody reactions. An actual experiment carried out along the schematic route presented in Figure 1 is shown in Figures 5, 6, 7, 8 and 9.

The genetic elements specified here and above may further and advantageously be equipped with at least two different selection markers functional in bacteria such as E.coli. Such selection markers, for example aphA (Pansegrau et al., 1987) or bla allow the easy separation of said genetic elements upon retransformation into E.coli strains.

In a preferred embodiment of the method of the present invention said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

Accordingly, the method of the invention is applicable in a wide range of biological interactions. For example, the invention will be useful in identifying peptide-protein or peptide-peptide-interactions by employing synthetic peptide libraries (Yang et al., 1995).

Two applications of interests are the application of a large scale two-hybrid system for the detection of protein-protein interactions involved in medically relevant pathways which may be useful as diagnostic or therapeutic targets for the treatment of disease, and a large scale tri-hybrid system which is one example of said complex of interacting molecules mentioned herein above for the identification of, for example, novel post-transciptional regulators and their binding sites (SenGupta et al., 1996; Putz et al., 1996). In this regard it should be noted that a complex, in accordance with the invention may comprise more than three interacting molecules. Furthermore, such a complex may be composed of biologically or chemically different members. For example, to identify interacting RNA binding proteins and RNA molecules, a plasmid expressing a LexA-HIV-1Rev protein, a plasmid transcribing an RNA sequence in fusion with the Rev responsive element and a plasmid expressing a potentially RNA-interacting protein in fusion with an activation domain may be present in one cell. The plasmids encoding the RNA fusion molecule and the activation domain fusion protein must contain different selectable and counterselectable markers according to the method of the invention. If the RNA

fusion molecule interacts with the respective two fusion proteins, the readout system is activated. To test whether the RNA fusion molecule or the activation domain fusion protein interact, the method of the invention is used to investigate the activation of the readout system in the absence of either of these fusion molecules.

In a further preferred embodiment, said genetic elements are plasmids, artificial chromosomes, viruses or other extrachromosomal elements.

Whereas it is preferred, due to the easy handling, to employ plasmids that specify the genetic elements in accordance with the present invention, the persons skilled in the art will be able to devise other systems that carry said genetic elements and that are identified above.

In an additional preferred embodiment, said readout system is a detectable protein. A number of readout systems are known in the art and may, if necessary, be adapted to be useful in the method of the invention.

Most preferably, said detectable protein is that encoded by the gene lacZ, HIS3, URA3, LYS2, sacB or HPRT, respectively. As is well known in the art, the expression of the ß-gal enzyme in yeast can be used for the formation of a detectable blue colony after incubation in X-Gal solution. Of course, the method of the invention is not restricted for use of only one readout system. On the contrary, if desired, a number of such readout systems may be combined. Said combination of a number of readout systems is, in accordance with the present invention, also comprised by the term "readout system". Such a combination will provide an additional safe guard for the identification of clones containing interacting partners.

Although the two-hybrid system has been developed in yeast, the method of the invention can be carried out in a variety of host systems. Preferred of those are yeast cells, bacterial cells, mammalian cells (Wu et al. 1996), insect cells or plant cells. Preferably, the bacterial cells are E. coli cells.

Of course, the genetic elements may be engineered and prepared in one host organism and then, e.g. by employing shuttle vectors, be transferred to a different host organism where it is employed in the method of the invention.

In another preferred embodiment, the method of the present invention comprises transforming or transfecting said host cell with at least one of said genetic elements prior to step (a).

Whereas the person skilled in the art may initiate the identification method of the invention starting from fully transformed or transfected host cells, he may wish to first generate such host cells in accordance with the aim of his research or commercial interest. For example, he may wish to generate a certain type of library first that he intends to screen against a second library already present in said host cells. Alternatively, he may have in mind to generate two or more different libraries that he wants to screen against each other. In this case, he would need to first transform said host cells, simultaneously or successively, with both or all types of genetic elements.

In another preferred embodiment, said host cells with said genetic elements are generated by cell fusion, conjugation or interaction mating.

The biological principal of counter-selection referred to above is well known in the art. Accordingly, the person skilled in the art may chose from a variety of such counter-selectable markers. Preferably, said markers are CAN1, CYH2, LYS2, URA3, HRPT or sacB.

It is further preferred in accordance with the present invention that said selectable markers are auxotrophic or antibiotic markers.

It is important to note that some of the markers that are used as a readout system, may also be used as selectable markers. It is further important to note that one and the same marker can not be used as selectable marker and as part of the readout system at the same time.

Most preferably, said auxotrophic or antibiotic markers are selected from LEU2, TRP1, URA3, HIS3, ADE2, LYS2 and Zeocin.

Planning of experiments may require that the test for interaction need not be done immediately after the provision of host cells and, possibly, the occurrence of the interactions. In such cases, the researcher may wish to store the transformed host cells for further use. Accordingly, a further preferred embodiment of the invention relates to a method wherein progeny of host cells obtained in step (b) are transferred to a storage compartment.

In particular in cases where a large number of clones is to be analyzed, said transfer is advantageously effected or assisted by automation or a picking robot. How such a picking robot may actually be put into practice, is described for example in Lehrach et al. (1997). Naturally, other automation or robot systems that reliably pick progeny of said host cells into predetermined arrays in the storage compartments may also be employed.

The host cells will, in this embodiment, be propagated in said storage compartment and provide further progeny for the additional tests. Preferably, replicas of said storage compartment maintaining the array of clones are set up. Said storage compartments comprising the transformed host cells and the appropriate media may be maintained in accordance with conventional cultivation protocols. Alternatively, said storage compartments may comprise an anti-freeze agent and therefore be appropriate for storage in a deep-freezer. This embodiment is particularly useful when the evaluation of potential interacting partners is to be postponed. As is well known in the art, frozen host cells may easily be recovered upon thawing and further tested in accordance with the invention. Most preferably, said anti-freeze agent is glycerol which is preferably present in said media in an amount of 3 - 25% (vol/vol).

In a further particularly preferred embodiment of the method of the invention, said storage compartment is a microtiter plate. Most preferably, said microtiter plate comprises 384 wells. Microtiter plates have the particular advantage of

providing a pre-fixed array that allows the easy replicating of clones and furthermore the unambiguous identification and assignment of clones throughout the various steps of the experiment. The 384 well microtiter plate is, due to its comparatively small size and large number of compartments, particularly suitable for experiments where large numbers of clones need to be screened.

Depending on the design of the experiment, the host cells may be grown in the storage compartment such as the above microtiter plate to logarithmic or stationary phase. Growth conditions may be established by the person skilled in the art according to conventional procedures. Cell growth is usually performed between 15 and 45 degrees Celsius.

Transfer of said host cells in step (c) is made or assisted by automation, by using a spotting robot or by using a pipetting or micropipetting device. How such a spotting robot may be devised and equipped is, for example, described in Lehrach et al. (1997). Naturally, other automation or robotic systems that reliably create ordered arrays of clones may also be employed.

Most preferably, said transfer is made to a planar carrier which is subsequently placed on the at least three selective media as specified in steps (ca) and (cb). Alternatively, said transfer of said host cells may be made to the planar carrier already placed on the selective media or said transfer may be made directly to the selective media.

Most advantageously, said transfer is effected in a regular grid pattern at densities of 1 to 1000 clones per square centimeter. The progeny of said host cells may be transferred to a variety of planar carriers. Most preferred is a membrane which may, for example, be manufactured from nylon, nitro-cellulose or PVDF.

The selective media used for growth of appropriate clones may be in liquid or in solid form. Preferably, said selective media when used in conjunction with a

spotting robot and membranes as planar carriers are solidified with agar on which said spotted membranes are subsequently placed. Alternatively, and also preferably, said selective media when in liquid form are held within microtiter plates and said transfer is made by replication.

Referring now to the step (d) of the method of the invention, the readout system can be analyzed by a variety of means. For example, it can be analyzed by visual inspection, radioactive, chemiluminescent, fluorescent, photometric, spectrometric, infra red, colourimetric or resonant detection.

Preferably, said identification of host cells that express interacting fusion proteins is effected by visual means from consideration of the activation state of said readout system of clones grown on the at least three selective media as specified in steps (ca) and (cb).

Also preferably, said identification of host cells that express interacting fusion proteins in step (d) is effected or assisted by digital image storage, analysis or processing. In this embodiment, positive clones which are preferably arrayed on a planar carrier such as a membrane are identified by comparison of digital images obtained from the membrane after activation of said readout system on said selective media specified in (ca) and (cb).

Most preferably, the identity of positive host cells and false positive host cells are stored on computer, for example within a relational database.

Identification of the at least one member of the pair or complex of interacting molecules may be effected by a variety of means. For example, molecules can be characterized by nucleic acid hybridization, oligonucleotide hybridization, nucleic acid or protein sequencing, restriction digestion, spectrometry or antibody reaction. Once the first member of an interaction has been identified, the second member or further members can also be identified by any of the above methods. Preferably the identification of at least one member of an

interaction is effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

If nucleic acid hybridization is to be carried out, the nucleic acid molecules comprised in the host cell and encoding for at least one of the interacting molecules is preferably affixed to a planar carrier. As is well known in the art, said planar carrier to which said nucleic acid may be affixed, can be for example, a Nylon-, nitrocellusose- or PVDF membrane, glass or silica substrates (DeRisi et al. 1996; Lockhart et al. 1996). Said host cells containing said nucleic acid may be transferred to said planar carrier and subsequently lysed on the carrier and the nucleic acid released by said lysis is affixed to the same position by appropriate treatment. Alternatively, progeny of the host cells may be lysed in a storage compartment and the crude or purified nucleic acid obtained is then transferred and subsequently affixed to said planar carrier. Advantageously, said nucleic acids are amplified by PCR prior to transfer to the planar carrier. Most preferably said nucleic acid is affixed in a regular grid pattern in parallel with additional nucleic acids representing different genetic elements encoding interacting molecules. As is well known in the art, such regular grid patterns may be at densities of between 1 and 50 000 elements per square centimeter and can be made by a variety of methods. Preferably, said regular patterns are constructed using automation or a spotting robot such as described in Lehrach et al. (1997) and Maier et al. (1997) and furnished with defined spotting patterns, barcode reading and data recording abilities. Thus it is possible to correctly and unambiguously return to stored host cells containing said nucleic acid from a given spotted position on the planar carrier. Also preferably, said regular grid patterns may be made by pipetting systems, or by microarraying technologies as described by Shalon et al. (1996), Schober et al (1993) or Lockart et al. (1996). Identification is, again, advantageously effected by nucleic acid hybridization.

Using a detectable nucleic acid probe of interest, homologous nucleic acids which are affixed on the planar carrier can be identified by hybridization. From the spotted position of said homologous identified nucleic acid on the planar

carrier, the corresponding host cell in the storage compartment can be identified which contains both or all members of the interaction. The for example second member of the interaction can now be identified by any of the above methods. For example, by use of a radioactively labeled Ras probe, homologous nucleic acids on the planar carrier can be identified by hybridization. The Ras interacting proteins can now be identified from the corresponding host cell that contains both the first genetic element homologous to the Ras probe and the second genetic element encoding for these Ras interacting proteins.

If multiple oligonucleotide hybridizations are carried out on the nucleic acids affixed to the planar carrier, oligofingerprints of all genetic elements encoding the interacting proteins can be obtained. These oligofingerprints can be used to identify all members of the interactions or those members that belong to specific gene families, as described in Maier et al. (1997).

Advantageously, the nucleic acid molecules that encode the interacting proteins are, prior to identification such as by DNA sequencing, amplified by PCR or in said genetic elements in host cells and preferable in *E. coli*. Amplification of said genetic elements is conducted by multiplication of the *E. coli* cells and isolation of said genetic elements. Methods of identifying the nucleic acids that encode interacting proteins by DNA sequencing and analysis are well known in the art. By amplifying and sequencing the nucleic acids that encode for both or all members of an interaction from the same clone, the identity of both or all members of the interaction can be determined.

If a specific antibody is to be used to determine whether a protein of interest is expressed as a fusion protein within an interaction library, it is advantageous to affix all fusion proteins expressed from the interaction library on to a planar carrier. For example, clones of the interaction library that express fusion proteins can be transferred to a planar carrier using a spotting robot as described in Lehrach et al (1997). The clones are subsequently lysed on the carrier and released proteins are affixed onto the same position. Using, for example, an anti-HIP1-antibody (Wanker et al. 1997), clones from the

interaction library that contain HIP1 fusion proteins and an unknown interacting fusion protein can be identified. The unknown member of the interacting pair of molecules can now be identified from the corresponding host cell by any of the above methods. The antibodies used as probes may be directly detectably labeled. Alternatively, said antibodies may be detected by a secondary probe or antibody which may be specific for the primary antibody. Various alternative embodiments using, for example, tertiary antibodies may be devised by the person skilled in the art on the basis of his common knowledge.

Most advantageously, when said identification of members comprising an interaction is effected using said regular grids, a digital image of the planar carrier after hybridization or antibody reaction is obtained and analysis is effected by digital image storage, processing or analysis using an automated or semi-automated image analysis system, such as described in Lehrach et al. (1997).

Most preferably, the information comprising the identity of the host cell and the identity of the interacting molecules expressed by the genetic elements contained within the host cell are stored on a computer, for example within a relational database.

These data are available for the establishment of a network of interactions. By collecting the information from a whole interaction library, the inter-relationship between many different interacting molecules can be determined and thus enable the establishment of a network of interactions. Preferably, said data can be accessed though the use of software tools or graphical interfaces that enable the investigator to easily query the established interaction network with a biological question or to develop the established network by the addition of further data.

Advantageously, those molecules identified as interacting with many different molecules can be recorded. This information can reduce the work needed to further characterize particular interactions since those interactions comprising

of a molecule found to interact with many other molecules within the yeast two-hybrid system may be suspected of being artifactual (Bartel et al., 1993).

A significant advantage of the method of invention over existing yeast two-hybrid systems is the scale at which such identification of interactions and interaction members can be made. Preferably, the method of invention screens library vs library interactions using arrayed interaction libraries. Thus, the method of invention allows, in an efficient manner, a more complete and exhaustive generation of protein-protein interaction networks than existing methods. An established and exhaustive network of protein-protein interactions is of use for many purposes as shown Figure 2. For example, it may be used to predict the existence of new biological interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localization of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins or interactions between proteins within a medically relevant pathway which are suitable for therapeutic intervention, diagnosis or the treatment of a disease.

In accordance with the present invention, it is additionally preferred prior to step (a) that a preselection against clones that express a single molecule able to activate the readout system is carried out on culture media comprising a counterselective compound, for example 5-fluoro orotic acid, canavanine, cycloheximide or α-amino-adipate .

In this embodiment, for example, the URA3 gene is incorporated as a component of the readout system. Clones containing only one of said genetic elements are placed on a selective medium comprising 5-fluoro orotic acid (5-FOA). In the case that clones that express a single molecule able to activate the readout system, 5-FOA is converted into the toxic 5-fluorouracil. Accordingly, host cells containing auto-activating molecules will die on the selective medium containing 5-FOA.

It is further important to note that the marker used for said preselection cannot be used as a selectable or counterselectable marker at the same time.

The present invention also relates to a method for the production of a pharmaceutical composition comprising formulation said at least one member of the interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

Said pharmaceutical composition comprises at least one of the aforementioned compounds identified by the method of the invention, either alone or in combination, and optionally a pharmaceutically acceptable carrier or exipient. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by conventional methods. These pharmaceutical compositions can be administered to subject in need thereof at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Dosages will vary but a preferred dosage for intravenous administration of DNA is from approximately 10^6 to 10^{22} copies of the nucleic acid molecule. Proteins or peptides may be administered in the range of 0,1ng to 10mg per kg of body weight. The compositions of the invention may be administered locally or systematically. Administration will generally be parenterally, e.g., intravenously; DNA may also be administered directly to the target site, e.g., by biolistic delivery to an internal or external target site or by catheter to a site in an artery.

The present invention further relates to a method for the production of a pharmaceutical composition comprising formulating an inhibitor of the interaction of the interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

The inhibitor may be identified according to conventional protocols. Additionally, molecules that inhibit existing protein-protein interactions can be isolated with the yeast two-hybrid system using the URA3 readout system. Yeast cells that express interacting GAL4ad and LexA fusion proteins which activate the URA3 readout system are unable to grow on selective medium containing 5-FOA. However, when an additional molecule is present in these cells which disrupts the interaction of the fusion proteins the URA3 readout system is not activated and the yeast cells can grow on selective medium containing 5-FOA. Using this method potential inhibitors of a protein-protein interaction can be isolated from a library comprising these inhibitors. Systems corresponding to the URA3 system may be devised by the person skilled in the art on the basis of the teachings of the present invention and are also comprised thereby.

Also, the present invention relates to a method for the production of a pharmaceutical composition comprising identifying a further molecule in a cascade of interacting molecules, of which the at least one member of interacting molecules identified by any of the above methods is a part of or identifying an inhibitor of said further molecule.

Once at least one member of the interacting molecules has been identified, it is reasonable to expect that said member is a part of a biological cascade. Identification of additional members of said cascade can be effected either by applying the method of the present invention or by applying conventional methods. Also, inhibitors of said further members can be identified and can be formulated into pharmaceutical compositions.

The present invention relates further to a kit comprising at least one of the following:

(i) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules containing a counterselectable marker and specified herein above;

(ii) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(iii) at least one genetic element comprising said genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(iv) at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(v) host cells comprising at least one and preferably at least two of said genetic elements specified in (iii) or (iv);

(vi) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified herein above wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(vii) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing genetic elements, host cells, storage compartments or carriers identified in any of (i) to (vi); and/or

(viii) at least one yeast strain comprising a *can*1 and a *cyh*2 mutation.

Preferably, said kit comprises or also comprises at least one storage compartment containing the host cells of (i), (ii) or (v) and/or comprises or also comprises at least one storage compartment containing said genetic information or said potentially interacting molecules encoded by said genetic information as specified in (i) or (iii).

The present invention also relates to the use of any of the yeast strains described herein above and in the appended examples for the the identification of at least one member of a pair of potentially interacting molecules.

The figures show:

Figure 1 A schematic overview of one embodiment of the method of the invention.

For the parallel analysis of a network of protein-protein interactions using the method of the invention, a library of plasmid constructs that express DNA binding domain and activation domain fusion proteins is provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into the improved binding domain and activating domain plasmids of the invention which contain different selectable and counterselectable markers. Both libraries are combined within yeast cells by transformation or interaction mating, and yeast strains that express potentially interacting proteins are selected on selective medium lacking histidine. The selective markers TRP1 and LEU2 maintain the plasmids in the yeast strains grown on selective media, whereas CAN1 and CYH2 specify the counter-selectable markers that select for the loss of each plasmid. HIS3 and lacZ represent selectable markers in the yeast genome, which are expressed upon activation by interacting fusion proteins. The readout system is, in the present case, both growth on medium lacking histidine and the enzymatic activity of ß-galactosidase which can be subsequently screened. A colony picking robot is used to pick the resulting yeast colonies into individual wells of 384-well microtiter plates containing selective medium lacking histidine, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library held in the microtiter plates

optionally may be replicated and stored. The interaction library is investigated to detect positive clones that express interacting fusion proteins and discriminate them from false positive clones using the method of the invention. Using a spotting robot, cells are transferred to replica membranes which are subsequently placed onto one of each of the selective media SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX. After incubation on the selective plates, the clones which have grown on the membranes are subjected to a ß-Gal assay and a digital image from each membrane is obtained with a CCD camera which is then stored on computer. Using digital image processing and analysis (Lehrach et al. 1997) clones that express interacting fusion proteins can be identified by considering the pattern of ß-Gal activity of these clones grown on the various selective media. The individual members comprising the interactions can then be identified by one or more techniques, including PCR, sequencing, hybridization, oligofingerprinting or antibody reactions.

Figure 2 The applications of an established and exhaustive network of protein-protein interactions. The identity of positive clones and the identity of the members comprising the interactions for the entire interaction library can be stored in a database. These data are used to establish a network of protein-protein interactions which can be used for a variety of purposes. For example, they may be used to predict the existence of new biological interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localization of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins within a medically relevant pathway which are suitable for

therapeutic, diagnosis intervention and for the treatment of disease.

Figure 3 Plasmids constructed for the improved 2-hybrid system.

a) the plasmid maps of the pBTM118a, b and c DNA binding domain vector series and the pGAD428a, b and c activation domain vector series. Both plasmids contain the unique restriction enzyme sites for *Sal* I and *Not* I which can be used to clone a genetic fragment into the multiple cloning site. The plasmids are maintained in yeast cells by the selectable markers TRP1 and LEU2 respectively. The loss of the plasmids can be selected for by the counterselective markers CAN1 and CYH2 respectively.

b) Polylinkers used within the multiple cloning site to provide expression of the genetic fragment in one of the three reading frames.

Figure 4 Identification of positive clones that contained interacting fusion proteins from false positive clones using the method of the invention.

Three different yeast clones each containing pairs of plasmid constructs (positive control: pBTM117c-SIM1 & pGAD426-ARNT; negative control: pBTM117c & pGAD426 and false-positive control: pBTM117c-HIP1 & pGAD426) were transferred by hand to four agar plates each containing a different selective medium (SD-leu-trp, SD-leu-trp-his, SD-leu+CAN and SD-trp+CAN), and incubated for 48 hours at 30 °C. The yeast colonies were subsequently transferred to a Nylon membrane and assayed for β-gal activity by the method of Breeden and Nasmyth (1985).

Figure 5 Predicted interactions between fusion proteins used to create the defined interaction library. The fusion proteins enclosed with dark rounded boxes are believed to interact as shown. The LexA-HIP1

fusion protein enclosed by a thin rectangular box has been shown to activate the LacZ readout system without the need for any interacting fusion protein. The two proteins LexA and GAL4ad, and the two fusion proteins GAL4ad-14-3-3 and LexA-MJD (all unboxed) are believed not to interact with each other or other fusion proteins used in this example.

Figure 6 Digital images of the ß-gal assays made from the replica Nylon membranes containing the spotted interaction library obtained from the selective media (a) SD-leu-trp-his, (b) SD-trp+CHX and (c) SD-leu+CAN.
In each case, The left hand side of each membrane contains control clones and clones from the defined interaction library, and the right hand side contains only clones from the defined interaction library. The two regions marked on the first membrane represent those clones magnified in Figure 7. The overall size of each membrane is 22 x 8 cm and contains 6912 spot locations at a spotting pitch of 1.4 mm.

Figure 7 Magnification of clones from the interaction library taken from the same region of three membranes obtained from the selective media SD-leu-trp-his, SD-trp+CHX and SD-leu+CAN assayed for β-gal activity:

A. Clones imaged from a region of the right hand side of the membrane containing the defined interaction library. Clones from the defined interaction library that express interacting proteins are ringed and correspond to the microtiter plate addresses 06L22 and 08N24.

B. Clones imaged from a region of the left hand side of the same membranes containing control clones and clones from the interaction library, where clones around each ink guide-spot are arranged as shown and correspond to: 00 Ink guide spot; 01False positive control clone that expresses the fusion protein GAL4ad-

LexA; 02 False positive clone expressing the fusion protein LexA-HIP1; 03 Positive control clone expressing the interacting fusion proteins LexA-SIM1 & GAL4ad-ARNT; 04 Clone from the defined interaction library. The positive control clone (spot position 03) is ringed.

Figure 8 Identification by hybridization of the genetic fragments carried by the clones 06L22 and 08N24. A 1.3 kb, SIM1 and a 1.4 kb ARNT DNA fragment were used as nucleic acid probes for hybridization to high-density spotted membranes containing DNA from the defined interaction library. These clones were identified containing SIM1 and ARNT genetic fragments by hybridization. The images are of the same region of the membranes as those shown in Figure 7a. The spot positions of the clones 06L22 and 08N24 are ringed.

Figure 9 Identification of the SIM1 and ARNT DNA fragments from the yeast two hybrid plasmid carried by the clone 06L22 by duplex PCR. Plasmid DNA was isolated from a liquid culture of the clone 06L22 by a QiaPrep (Hilden) procedure and the inserts contained within the plasmids were amplified by PCR using the primer pairs, 5'-TCG TAG ATC TTC GTC AGC AG-3' & 5'-GGA ATT AGC TTG GCT GCA GC-3' for the plasmid pBTM117c and 5'-CGA TGA TGA AGA TAC CCC AC-3' & 5'-GCA CAG TTG AAG TGA ACT TGC-3' for pGAD426. Lane 1 contains a Lamda DNA digestion with *Bst*EII as size marker; Lane 2 contains the duplex PCR reaction from plasmids isolated from clone 06L22; Lanes 3 and contain control PCR amplifications from the plasmids pBTM117c-SIM1 and pGAD426-ARNT respectively.

The examples illustrates the invention.

Example 1. Construction of vectors and a novel host strain for an improved yeast two-hybrid system

The plasmids constructed for an improved yeast two-hybrid system pBTM118 a, b and c and pGAD428 a, b and c are shown in Fig. 3a. Both sets of vectors can be used for the construction of hybrid (fusion) proteins. The vectors contain the unique restriction sites *Sal* I and *Not* I located in the multiple cloning site (MCS) region at the 3´- end of the open reading frame for either the *lexA* coding sequence or the GAL4ad sequence (Fig. 3b).

With both sets of plasmids fusion proteins are expressed at high levels in yeast host cells from the constitutive *ADH1* promoter (P) and the transcription is terminated at the *ADH1* transcription termination signal (T). The two-hybrid plasmids shown in Fig. 3a are shuttle vectors that replicate autonomously in both *E. coli* and *S. cerevisiae*.

The three plasmids pBTM118 a, b and c are used to generate fusions of the LexA protein (amino acids 1-220) and a protein of interest cloned into the MCS in the correct orientation and reading frame. The plasmids pBTM118 a, b and c are derived from pBTM117c (Wanker et al., 1997) by insertion of the adapters shown in Table 1 into the restriction sites *Sal* I and *Not* I to generate the improved vectors with three different reading frames.

The plasmids pBTM118 a, b and c carry the wild type yeast *CAN1* gene for counterselection, which confers sensitivity to canavanine in transformed yeast cells (Hoffmann, 1985). The plasmids also contain the selectable marker *TRP1*, that allows yeast trp1-auxotrophs to grow on selective synthetic medium without tryptophan, and the selectable marker *bla* which confers ampicillin resistance in *E. coli*.

The plasmids pGAD428 a, b and c are used to generate fusion proteins that contain the GAL4 activation domain (amino acids 768-881) operatively linked to a protein of interest. The plasmids pGAD428 a, b and c carry the wild type

yeast *CYH2* gene, which confers sensitivity to cycloheximide in transformed cells (Kaeufer et al., 1983), the selectable marker *LEU2*, that allows yeast leu2- auxotrophs to grow on selective synthetic medium without leucine, and the bacterial marker *aphA* (Pansegrau et al., 1987) which confers kanamycin resistance in *E. coli*. The plasmids pGAD428a, b and c were created from pGAD427 by ligation of the adapters shown in Table 1 into the MCS to construct the improved vectors with three different reading frames.

For the construction of pGAD427 a 1.2 kb *Dde* I fragment containing the *aphA* gene was isolated from pFG101u (Pansegrau et al., 1987) and was subcloned into the *Pvu* I site of the pGAD426 using the oligonucleotide adapters 5'-GTCGCGATC-3' and 5'-TAAGATCGCGACAT-3'. The plasmid pGAD426 was generated by insertion of a 1.2 kb *Eco* RV *CYH2* gene fragment, which was isolated from the pAS2-1 (Clonetech) into the *Pvu* II site of pGAD425 (Han and Collicelli, 1995).

Table 1: Oligonucleotide adapters for the construction of the novel yeast two-hybrid vectors pBTM118 a, b and c and pGAD428 a, b and c.

oligonucleotide	sequence (5'-3')
a sense	TCGAGTCGACGCGGCCGCTAA
a antisense	GGCCTTAGCGGCCGCGTCGAC
b sense	TCGAGGTCGACGCGGCCGCAGTAA
b antisense	GGCCTTACTGCGGCCGCGTCGACC
c sense	TCGAGAGTCGACGCGGCCGCTTAA
c antisense	GGCCTTAAGCGGCCGCGTCGACTC

To allow for the double counterselection provided by the improved two-hybrid vectors, the *S. cerevisiae* strain L40cc was created. L40cc is isogenic with strain L40c (Wanker *et al,* 1997.), except for the presence of a cyh2 mutation. This mutation was selected by plating L40c cells onto YPD plates containing 10

μg/ml cycloheximide (Sigma St Louis). The genotype of the L40cc strain is: *Mata his3Δ200 trp1-910 leu2-3,112 ade2 LYS2::(lexAop)$_4$-HIS3 URA3::(lexAop)$_8$-lacZ Gal4 can1 cyh2*.

Example 2. Detection of clones expressing known interacting proteins from false positives using the improved two-hybrid system

Any large scale application of the yeast two-hybrid system requires a reliable method for the detection of real positive clones containing interacting proteins from false positive clones. In particular, to detect those false positive clones that express fusion proteins, which are able to activate the readout system without any interaction with a second molecule it is necessary to analyze also cells expressing only the single fusion proteins. To therefore facilitate the selective elimination of certain plasmids encoding one of each fusion protein we integrated two independent countersection features into the two-hybrid system. We generated a yeast strain, L40cc, which is resistent against canavanine and cycloheximide and constructed the vectors pBTM118 a, b and c and pGAD428 a, b and c containing the *CAN1* and the *CYH2* wild type gene, respectively. After transforming both plasmids into this strain and plating the cells onto selective media containing either canavanine or cycloheximide, only those cells within the cell population which have lost the plasmid expressing the respective wild type gene can grow. The cells that grow therefore contain only the remaining plasmid which encodes a single fusion protein and can easily be analyzed for its ability to activate lacZ expression by a subsequent ß-galactosidase assay.

To determine whether the two-hybrid plasmids described in Example 1 can be used for the detection of clones expressing interacting molecules from false positive clones several DNA fragments encoding proteins of interest were cloned into the vectors. The orientation of the inserted fragments was

determined by restriction analysis and the reading frame was checked by sequencing. The generated constructs and the original plasmids described above are listed in Table 2. The construction of pBTM117c-HD1.6, -HD3.6 and -SIM1 was described elsewhere (Wanker et al., 1997; Probst et al., 1997). pBTM117c-HIP1 and pGAD426-HIP1 were obtained by ligation of a 1.2 kb *Sal* I HIP1 fragment isolated from pGAD-HIP1 (Wanker et al., 1997) into pBTM117c and pGAD426, respectively. pBTM117c-MJD was created by inserting a 1.1 kb *Sal* I/*Not* I MJD1 fragment (Kawagushi et al., 1994) into pBTM117c, and pGAD426-14-3-3 was generated by inserting a 1.0 kb EcoRI/NotI fragment of pGAD10-14-3-3 into pGAD426. For the construction of pGAD426-HIPCT, a 0.5 kb *Eco* RI HIP1 fragment isolated from pGAD10-HIPCT was ligated into pGAD426. pLEXA-HIP1 and pGAD426-ARNT were generated by inserting a 2.5 kb *Sph* I LexA-HIP1 fragment and a 1.4 kb *Sal* I/*Not* I ARNT fragment into pGAD426, respectively.

It was shown that the fusion proteins LexA-SIM1 and GAL4ad-ARNT specifically interact with each other in the yeast two-hybrid system (Probst et al., 1997), because when both hybrids were coexpressed in *Saccharomyces cerevisiae* containing two integrated reporter constructs, the yeast HIS3 gene and the bacterial lacZ gene, which both contained binding sites for the LexA protein in the promoter region, the interaction between these two fusion proteins led to the transcription of the reporter genes. The fusion proteins by themselves were not able to activate transcription because GAL4ad-ARNT lacks a DNA binding domain and LexA-SIM1 an activation domain (Probst et al., 1997). In contrast it was shown recently that the fusion protein LexA-HIP1 is capable of activating the HIS3 and lacZ reporter genes without interacting with a specific GAL4ad fusion protein. Thus, the yeast clones expressing the LexA-HIP1 protein have to be designated as false positives, because false positives are defined here as clones where a GAL4ad fusion protein or a LexA fusion protein alone without the respective partner protein activates the transcription of the reporter genes without the need for any interacting partner protein. To differentiate between positive clones that express interacting fusion proteins and false positives, an improved version of the two hybrid system described in

this invention was developed. In this improved method the plasmids for the expression of the LexA and GAL4ad fusion proteins contain in addition to the selectable markers *TRP1* and *LEU2* also the counterselectable markers *CAN1* and *CHY2*, respectively (Table 2). When both plasmids are transformed into the yeast strain L40cc which is cycloheximide (*cyh2*) and canavanine (*can1*) resistant, the resulting transformants become sensitive to cycloheximide and canavanine (Hoffman, 1985; Kaeufer et al., 1983). Thus, when the L40cc yeast strain expressing the fusion proteins LexA-SIM1 and GAL4ad-ARNT is selected on canavanine medium without leucine (SD-leu+CAN), only those cells of phenotype Leu+CAN- can grow, will maintain the LEU2 plasmid (pGAD426-ARNT) and should show a LacZ- phenotype, because the TRP1/CAN1 plasmid encoding the LexA-SIM1 fusion protein has been lost through counterselection. The same result should also be obtained when this strain is selected on cycloheximide medium without tryptophan (SD-trp+CHX), because on this medium only yeast cells which have lost the LEU2/CYH2 plasmid through counterselection and maintain the TRP1 plasmid (pBTM117c-SIM1) can grow. In contrast, when the L40cc strain expressing the LexA-HIP1 fusion protein encoded by the pBTM117c-HIP1 is selected on SD-leu+CAN and SD-trp+CHX medium the resulting strains should have the respective phenotype lacZ- and lacZ+. Thus, with this improved two hybrid method the discrimination of positive clones that express interacting proteins from false positives should be possible by selecting the yeast cells in selective medium lacking histidine (SD-leu-trp-his) and in parallel selecting them on SD-leu+CAN and SD-trp+CHX medium. To test this hypothesis, pairs of the yeast two-hybrid plasmids pBTM117cSIM1 & pGAD426-ARNT, pBTM117c & pGAD426 and pBTM117c-HIP1 & pGAD426 were transformed into the yeast strain L40cc and Trp+Leu+ transformants that contained at least one of each of the two plasmids were selected on SD-leu-trp plates. Two transformants from each transformation were investigated for the presence of protein-protein interactions by testing the ability of the yeast cells to grow on the SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX plates and by the ß-galactosidase membrane assay (Breeden and Nasmyth, 1985). Fig. 4 shows that the yeast strains cells harboring both the plasmids pBTM117c-SIM1 & GAD426-ARNT or pBTM117c-HIP1 & pGAD426 grow on SD-leu-trp-his plates

and turned blue after incubation in X-Gal solution, indicating that the HIS3 and lacZ reporter genes are activated in these strains. In comparison, the yeast strain harboring both the negative control plasmids pBTM117c & pGAD426 was not able to grow on this medium and also showed no lacZ activity. After selection of the yeast strains harboring the different combinations of the two-hybrid plasmids on SD-leu+CAN and SD-trp+CHX the resulting strains were also analyzed by the ß-galactosidase assay. After incubating the membrane containing all three strains on SD-trp+CHX medium only progeny of the yeast strain that originally harbored both the plasmids pBTM117c-HIP1 & pGAD426 yet which had lost the pGAD426 plasmid through counterselection turned blue after incubating in X-Gal solution. This result indicates that this clone is a false positive, because although showing a lacZ+ phenotype when grown on SD-leu-trp-his medium, the LexA-HIP1 fusion protein was also capable of activating the HIS3 and lacZ genes on SD-trp+CAN medium without the need for any interacting fusion protein. In comparison, the yeast strain harboring both the plasmids pBTM117c-SIM1 & pGAD426-ARNT is a positive clone that expresses interacting LexA and GAL4ad fusion proteins, because both the LexA and the Gal4ad fusion proteins are necessary for the activation of the reporter genes. If either of the plasmids pBTM117c-SIM1 or pGAD426-ARNT is lost from the strain by counterselection on SD-trp+CHX or SD-leu+CAN, respectively, the resulting cells are no longer able to activate the lacZ reporter gene and do not turn blue after incubation in X-GAL solution. With the membranes from the SD-leu+CAN plate false positive clones expressing an autoactivating GAL4ad-LexA fusion protein were also detected by the ß-galactosidase assay.

Table 2: Two-hybrid vectors used for the expression of fusion proteins.

plasmid	fusionprotein	insert (kb)	counterselection	selection in yeast	reference
pBTM117c	lexA	-	CAN1	TRP1	Wanker et al. 1997
pBTM117c-HD1.6	lexA-HD1.6	1.6	CAN1	TRP1	Wanker et al., 1997
pBTM117c-HD3.6	lexA-HD3.6	3.6	CAN1	TRP1	Wanker et al., 1997
pBTM117c-SIM1	lexA-SIM1	1.1	CAN1	TRP1	Probst et al., 1997
pBTM117c-MJD	lexA-MJD	1.4	CAN1	TRP1	this work
pBTM117c-HIP1	lexA-HIP1	1.2	CAN1	TRP1	this work
pGAD426	GAL4ad	-	CYH2	LEU2	this work
pGAD426-ARNT	GAL4ad-ARNT	1.3	CYH2	LEU2	Probst et al., 1997
pGAD426-HIP1	GAL4ad-HIP1	1.2	CYH2	LEU2	Wanker et al., 1997
pGAD426-HIPCT	GAL4ad-HIPCT	0.8	CYH2	LEU2	Wanker et al., 1997
pGAD426-14-3-3	GAL4ad-14-3-3	1.0	CYH2	LEU2	this work
pLEXA-HIP1	lexA-HIP1	1.2	CYH2	LEU2	this work

Example 3. Detection and identification of interacting proteins using a large-scale and automated application of the improved 2-hybrid system.

A scheme utilizing the method of the invention within a large-scale and automated approach for the parallel detection of clones that express interacting fusion proteins and the identification of members comprising the interactions is shown in Figure 1. Yeast clones from an 'interaction library' that express interacting proteins are identified on a large-scale by the use of visual inspection or digital image processing and analysis of high-density spotted membranes on which their ß-galactosidase activity has been assayed after growth on various selective media. Automated methods based on those described in Lehrach *et al.* (1997) are used to effect the production of the interaction library and high-density spotted membranes, and the analysis of digital images of the ß-gal assay and hybridization images.

To prove that the method of the invention as described in Figure 1 could successfully identify positive clones that expressed interacting proteins from false positive clones, and then subsequently identify the individual members comprising the interaction, an experiment was conducted using well defined plasmid constructs for the expression of known fusion proteins. Some of these fusion proteins are known to interact with each other while others do not interact with any other fusion proteins in the defined system. The essential steps of the method shown in Figure 1 were used, and the results show that the method of the invention can be used as a high-throughput, parallel and automated approach to generate large amounts of data leading to the establishment of protein-protein interacting networks.

Generation of a well defined interaction library

To generate the well defined interaction library, a series of plasmid constructs were used. Table 2 lists the constructs used for the expression of the LexA or GAL4ad fusion proteins. The predicted protein-protein interactions of these fusion proteins are shown in Figure 5. It was shown that the fusion proteins LexA-SIM1 & GAL4ad-ARNT and LexA-HD1.6 & GAL4ad-HIP1 specifically interact with each other in the yeast two-hybrid system because they only activate the reporter genes HIS3 and LacZ when both proteins are present in one cell (Probst et al. 1997; Wanker et al. 1997). In contrast, it was demonstrated that the LexA-HIP1 fusion protein is capable of activating the reporter genes without the need for any interacting fusion protein. The proteins LexA and GAL4ad and the fusion proteins LexA-MJD and GAL4ad-14-3-3 which are also present in the defined interaction library are unable to activate the reporter genes either alone or when present in the same cell with any other fusion proteins comprising the library.

To generate the well defined interaction library, the constructs for the expression of the nine fusion proteins shown in Figure 5 were pooled and 3 µg of the mixture was co-transformed into yeast strain L40cc by the method of Schiestel & Gietz (1989). The resulting transformants were plated onto large 24

x 24 cm agar plates (Genetix, UK) containing minimal medium lacking tryptophan, leucine and histidine (SD-leu-trp-his). After growth at 30°C for 4 days, individual yeast colonies were picked using a picking robot based on that described in Lehrach *et al.* (1997). With this robot, individual yeast colonies were picked into individual wells of a 384-well microtiter plates (Genetix, UK) containing SD-leu-trp-his/7% glycerol liquid medium. The resulting microtiter plates were incubated at 30°C for 3 days. Although yeast colonies are more difficult than *E. coli* cells to handle in automated systems, a picking success of approximately 80% was achieved. After growth of yeast strains within the microtiter plates, each plate was labeled with an individual number and barcode. Each plate was also replicated to create two additional copies using a sterile 384-pin plastic replicator (Genetix, UK) to transfer a small amount of cell material from each well into pre-labeled 384-well microtiter plates and pre-filled with SD-leu-trp-his/7% glycerol liquid medium. The replicated plates were incubated at 30 °C for 3 days, subsequently frozen and stored at -70 °C together with the original picked microtiter plates of the interaction library.

Generation of high-density spotted membranes for use in an improved yeast 2-hybrid approach

A high-throughput spotting robot such as that described by Lehrach *et al.* (1997) was used to construct filters with a high-density pattern of yeast clones from the defined interaction library contained within 384-well microtiter plates. The position of individual clones on the high-density filter was recorded by the robot by the use of a pre-defined duplicate spotting pattern and the barcode of the microtiter plate. Labeled membranes (Hybond N+, Amersham UK) were pre-soaked in SD-leu-trp-his medium and placed in the robot. The interaction library was automatically arrayed as replica copies onto the membranes using a 384-pin spotting tool affixed to the robot. Five different microtiter plates from the first copy of the interaction library were replica spotted in a '3x3 duplicate' pattern around a central ink guide-spot onto 10 nylon membranes - corresponding to approximately 1900 clones spotted at a density of approximately 35 spots per cm^2. On each replica membrane three different control clones were spotted, each from a microtiter plate that contained the

same control clone in every well. One control clone expressed the fusion proteins LexA-SIM1 & GAL4ad-ARNT, a second control clone the fusion protein LexA-HIP1, while a third expressed fusion protein GAL4ad-LexA, and all were spotted in order to test the selection, counterselection and the ß-gal assay features of the method. To ensure the number of yeast cells on each spot was sufficient for those membranes which were to be placed on the counterselection media plates, the robot was programmed to spot onto each spot position 5 times from a slightly different position within the wells of the microtiter plates. The robot created a data-file in which the spotting pattern produced and the barcode that had been automatically read from each microtiter plate was recorded.

Each membrane was carefully laid onto approximately 300 ml of solid agar media in 24 x 24 cm assay trays. Six membranes were transferred to SD-leu-trp-his media and two each of the remaining membranes were transferred to either SD-trp+CHX or SD-leu+CAN media. The yeast colonies were allowed to grow on the surface of the membrane by incubation at 30 °C for 3 days.

Detection of the readout system

Two membranes from each of the selective media were assayed for lacZ expression using the ß-gal assay as described by Breeden & Nasmyth (1985) and air dried overnight. For each membrane, a 32-bit digital image of the ß-gal assay was obtained with a high-resolution charge coupled device (CCD) color camera (Kontron, Germany), and the images were stored on computer. One image of the defined interaction library that was grown on membranes placed on each of the 3 selective media and subsequently assayed for β-gal activity is shown in Figure 6. Individual clones of the interaction library can be identified and their position on the high-density spotted filter converted to specific wells in the microtiter plates using a semi-automated screening system as described by Lehrach *et al.* (1997).

Positive clones that express interacting fusion proteins can be detected from false positive clones by considering the activity of ß-galactosidase of clones

grown on spotted membranes laid on the various selective media. Positive clones should activate the lacZ reporter gene on SD-leu-trp-his media and turn blue on incubation with X-Gal solution, but not on either of the two counterselective media. False positive clones should activate the reporter gene and turn blue on incubation with X-Gal solution on at least one counterselective media as well as on the SD-leu-trp-his medium.

Figure 7 shows magnified images of a ß-gal assay of clones grown on the membranes which had been placed on the three selective media. Within the magnified region of the membranes shown in Figure 7 a, two clones were detected as positive clones that express interacting fusion proteins since they activated the lacZ reporter gene on SD-leu-trp-his media, but not on either of the two counterselective media, and whose spotted positions are circled. The two clones were identified by their microtiter plate address within the interaction library as 06L22 and 08N24 respectively. All other clones spotted within this region of the membrane were detected as false positive since they express ß-galactosidase on SD-trp+CHX medium as well as on SD-leu-trp-his medium.

Expression of the LacZ reporter gene for the three control clones spotted onto the same membranes confirm these results. The positive control clone that expresses the interacting fusion proteins LexA-SIM1 & GAL4ad-ARNT should show a LacZ+ phenotype when grown on SD-leu-trp-his medium, but LacZ- when grown on either of the counterselective media. This control clone was spotted at position 03 in the region of the membranes shown in Figure 7 b, of which one example is circled. The pattern of ß-gal activity for this positive control clone on the three selective media is as predicted. The false positive control clone that expresses the fusion protein LexA-HIP1 and the false positive clone that expresses the fusion protein GAL4ad-LexA are spotted at positions 02 and 01 respectively. Both false positive control clones show a LacZ+ phenotype when grown on SD-leu-trp-his media, but are detected as false positive clones by the method of the invention since they also show a LacZ+ phenotype on SD-leu+CAN or SD-trp+CHX media, respectively. The clones spotted at position 04 are from the defined interaction library, and from their

LacZ+ phenotype when grown on SD-leu+CAN media are predicted to be false positive clones.

Identification of individual members of the interaction

The interaction library constructed for this example was composed of known fusion proteins with predicted interactions as shown in Figure 5. A real positive clone from this defined interaction library is therefore expected to express the interacting fusion protein-pairs LexA-SIM1 & GAL4ad-ARNT or LexA-HD1.6 & GAL4ad-HIP1 and hence contain the corresponding pairs of plasmid constructs pBTM117c-SIM1 & pGAD426-ARNT or pBTM117c-HD1.6 & pGAD426-HIP1, respectively. The identification of individual members that comprise an interaction between fusion proteins that are expressed within a single cell can be made by a variety of means as outlined in Figures 1. and 2. Two independent methods, nucleic acid hybridization and PCR, were used to identify the individual plasmid constructs that expressed the interacting fusion proteins in the positive clones 06L22 and 08N24.

The four membranes which had been placed on the SD-leu-trp-his medium and had not been used to assay ß-gal activity were processed according to the procedure described in Larin & Lehrach (1990) in order to affix the DNA contained within the clones of the interaction library onto the surface of the membrane. A 1.3 kb DNA fragment of SIM1 and a 1.4 kb DNA fragment of ARNT were radioactively labeled by standard random priming procedures for use as a hybridization probe (Feinberg & Vogelstein, 1983). Each probe was heat denatured for 10 min at 95 °C and hybridized overnight at 65 °C in 15 ml of 5% SDS/0.5M sodium phosphate (pH 7.2)/1 mM EDTA with a high-density spotted membrane with DNA from the interaction library affixed to it. The membranes were washed once in 40mM sodium phosphate/0.1%SDS for 20 min at room temperature and once for 20 min at 65 °C before wrapping each membrane in Saran wrap and exposing it overnight to a phosphor-storage screens (Molecular Dynamics, USA). A digital image of each hybridized membrane was obtained by scanning the phosphor-storage screen using a phosphor-imager (Molecular Dynamics, USA). The digital image was stored on

computer and was analyzed using a semi-automated system as described in Lehrach *et al.* (1997) which marked positive hybridization signals with square blocks. Figure 8 shows a magnified region of each hybridized membrane corresponding to that shown in Figure 7 a containing the clones 06L22 and 08N24, the spotting position of which are circled. These clones were predicted to express either the interacting fusion protein-pairs LexA-SIM1 & GAL4adARNT or LexA-HD1.6 & GAL4ad-HIP1, and hybridization with the specific SIM1 and ARNT probes have shown that both clones contain the plasmid constructs pBTN117c-SIM1 and pGAD426-ARNT.

The individual clone 06L22 was recovered from the frozen plates of the original interaction library and inoculated into SD-leu-trp-his liquid medium. This culture was allowed to grow for 3 days at 30 °C and the corresponding plasmids contained in the clone were isolated using a QiaPrep (Qiagen, Hilden) procedure. Duplex PCR was used to simultaneously amplify the inserts contained within the plasmid constructs using primer-pairs specific for either the pBTM117c or GAD426 plasmids. The presence of the SIM1 and ARNT inserts was confirmed for clone 06L22 by electrophoresis of the amplified PCR products against separate control amplifications of the inserts from plasmids pBTM117c-SIM1 and pGAD426-ARNT as size markers (Figure 9).

References

Bartel, P., Chien, C.-T., Sternglanz, R. and Fields, S. (1993). Elimination of false positives that arise in using in the two-hybrid system. *BioTechniques* **14**: 920-924

Bartel, P.L., Roecklein, J.A., SenGupta, D., and Fields, S. (1996). A protein linkage map of
Escherichia coli bacteriophage T7. *Nat. Genet.* **12**: 72-77.

Breeden, L. and K. Nasmyth, K. (1985). Regulation of yeast HO gene. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 643-650.

Cubbitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**: 448-455.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**: 457-460.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specifity activity. *Anal. Biochem.* **132**: 6-13

Fields, S.and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Han, L. and Colicelli, J. (1995). A human protein selected for interference with Ras function interacts directly with Ras and competes with Raf1. *Mol. Cell. Biol.* **15**: 1318-1323.

Hoffmann, W. (1985). Molecular characterization of the CAN1 locus in Saccharomyces cerevisiae. *J. Biol. Chem.* **260**: 11831-11837.

Kaeufer, N.F., Fried, H.M., Schwindinger, W.F., Jasin, M. and Warner, J.R. (1983). Cycloheximide resistance in yeast: the gene and its protein. *Nucleic Acids Res.* **11**: 3123-3135.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S. and Kakizuka, A. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **8**: 221-228.

Larin, Z. and Lehrach, H. (1990). Yeast artificial chromosomes: an alternative approach to the molecular analysis of mouse developmental mutations. *Genet. Res.* **56**: 203-208.

Lehrach, H., Bancroft, D. and Maier, E. (1997). Robotics, computing, and biology: An interdisciplinary approach to the analysis of complex genomes. *Interdisp. Science Rev.* **22**: 37-43.

Li, J.J. and Herskowitz, I. (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* **262**:1870-1874.

Lockhart, D. J., Dong, H. Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. nature Biotechnology 12: 1675-1680.

Maier, E., Meier-Ewert, S., Bancroft, D. and Lehrach H. (1997). Automated array technology for gene expression profile. *Drug Disc. Today* **2**: 315-324.

Pansegrau, W., Miele, L., Lurz, R. and Lanka E. (1987). Nucleotide sequence of the kanamycin resistance determinant of plasmid RP4: homology to other aminoglycoside 3'-phosphotransferases. *Plasmid* **18**: 193-204.

Probst, M.R., Fan, C.M., Tessier-Lavigne, M. and Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* **272**: 4451-4457.

Putz, U., Skehel, P. and Kuhl, D. (1996). A tri-hybrid system for the analysis and detection of RNA--protein interactions. *Nucleic Acids Res.* **24**: 4838-4840 .

Schiestl, R.H. and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**: 339-346.

Schober, A., Guenther, R. Schwienhorst, A., Doering, M. and Lindemann, B. F. (1993). Accurate high-speed liquid handling of very small biological samples. Biotechniques 15: 324-329.

SenGupta DJ., Zhang B., Kreamer B., Pochart P., Fields S., Wickens M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *PNAS.* 93:8496-501

Shalon, D., Smith, S. J. and Brown, P. O. (1996). A DNA microarray system for analysing complex DNA samples using two-colour fluorescent probe hybridization. Geneome Research 6, 639-645.

Yang, M., Wu, Z. and Fields (1995). Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* **23**:1152-1156

Wanker, E.E., Rovira, C., Scherzinger, E., Hasenbank, R., Waelter, S., Tait, D., Colicelli, J. and Lehrach H. (1997). HIP-I: a huntingtin interacting protein isolated by the yeast two-hybrid system. *Hum. Mol. Genet.* **6**: 487-495.

CLAIMS

1. A method for the identification of at least one member of a pair or complex of interacting molecules, comprising:

   (a) providing host cells containing at least two genetic elements with different selectable and counter-selectable markers, said genetic elements each comprising genetic information specifying one of said members, said host cells further carrying a readout system that is activated upon the interaction of said molecules;

   (b) allowing at least one interaction, if any, to occur;

   (c) selecting for said interaction by transferring progeny of said host cells to:

      (ca) at least two different selective media, wherein each of said selective media allows growth of said host cells only in the absence of at least one of said counter-selectable markers and in the presence of a selectable marker; and

      (cb) a further selective medium that allows identification of said host cells only on the activation of said readout system;

   (d) identifying host cells containing interacting molecules that:

      (da) do not activate said readout system on any of said selective media specified in (ca); and

      (db) activate the readout system on said selective medium specified in (cb); and

   (e) identifying at least one member of said pair or complex of interacting molecules.

2. The method of claim 1 wherein said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., Xu, X.L., Yang, M.C., Hwang, L.Y., Bowcock, A.M. and Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat. Genet.* **14**: 430-440.

Zhang, J. and Lautar, S. (1996). A yeast three-hybrid method to clone ternary protein complex components. *Anal. Biochem.* **242**:68-72.

3. The method of claim 1 or 2 wherein said genetic elements are plasmids artificial chromosomes, viruses or other extra chromosomal elements.

4. The method of any one of claims 1 to 3 wherein said interactions lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety driving the activation of said readout system.

5. The method of claim 4 wherein said readout system is a detectable protein.

6. The method of claim 5 wherein said detectable protein is encoded from at least one of the genes lacZ, HIS3, URA3, LYS2, sacB or HRPT.

7. The method of any one of claims 1 to 6 wherein said host cells are yeast cells, bacterial cells, mammalian cells, insect cells or plant cells.

8. The method of any one of claims 1 to 7 further comprising transforming or transfecting said host cells with said genetic elements prior to step (a).

9. The method of any one of claims 1 to 8 wherein cell fusion, conjugation or interaction mating is used for the generation of said host cells with said genetic elements prior to step (a).

10. The method of any one of claims 1 to 9 wherein said counter-selectable markers selected against in step (ca) are selected from the group of CAN1, CYH2, LYS2, URA3, HPRT and sacB.

11. The method of any one of claims 1 to 10 wherein said selectable marker is an auxotrophic or antibiotic marker.

12. The method of claim 11 wherein said auxotrophic or antibiotic marker is LEU2, TRP1, URA3, ADE2, HIS3, LYS2 or Zeocin.

13. The method of any one of claims 1 to 12 wherein progeny of host cells of step (b) are transferred to storage compartment.

14. The method of claim 13 wherein said transfer is effected or assisted by automation or a picking robot.

15. The method of claim 13 or 14 wherein said storage compartment comprises an anti-freeze agent.

16. The method of any one of claims 3 to 15 wherein said storage compartment is a microtiter plate.

17. The method of claim 16 wherein said microtiter plate comprises 384 wells.

18. The method of any one of claims 1 to 17 wherein said transfer in step (c) is made or assisted by automation, a spotting robot, pipetting or micropipetting device.

19. The method of claim 18 wherein said transfer is made to a planar carrier.

20. The method of claim 18 or 19 wherein said transfer is in a regular grid pattern of densities of 1 to 1000 clones per cm^2.

21. The method of any one of claims 18 to 20 wherein said planar carrier is a membrane.

22. The method of any one of claims 1 to 21 wherein said identification of said host cells in step (d) is effected by visual means from consideration of the activation state of said readout system.

23. The method of any one of claims 1 to 22 wherein said identification of said host cells in step (d) is effected by digital image storage, analysis or processing.

24. The method of any one of claims 1 to 23 wherein said identification of said at least one member of said pair of interacting molecules is effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

25. The method of claim 24 wherein said identification made by said antibody reaction or said hybridization is effected using regular grids of said at least one member or of said genetic information encoding said at least one member.

26. The method of claim 25 wherein construction of said regular grids is effected by automation or a spotting robot.

27. The method of any one of claims 24 to 26 wherein said identification is effected by digital image storage, processing or analysis.

28. The method of any one of claims 24 to 27 wherein nucleic acid molecules, prior to said identification, are amplified by PCR or are amplified in as a part of said genetic elements, preferably in bacteria and most preferably in E.coli.

29. The method of any one of claims 1 to 28 wherein, prior to step (a) a preselection against clones that express a single molecule able to activate the readout system is carried out on culture media comprising a counterselective compound.

30. The method of claim 29 wherein said counterselective compound is 5-fluoro orotic acid, canavanine, cycloheximide or α-amino-adipate.

31. A method for the production of a pharmaceutical composition comprising formulating said at least one member of the interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

32. A method for the production of a pharmaceutical composition comprising formulating an inhibitor of the interaction of the interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

33. A method for the production of a pharmaceutical composition comprising identifying a further molecule of a cascade of interacting molecules, of which the at least one member of said interacting molecules identified by the method of any one of claims 1 to 30 is a part of or identifying an inhibitor of said further molecule.

34. Kit comprising at least one of the following:

(i) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(ii) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(iii) at least one genetic element comprising said genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(iv) at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules

containing a counterselectable marker and specified in any of the preceding claims;

(v) host cells comprising at least one and preferably at least two of said genetic elements specified in (iii) or (iv);

(vi) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified in any of the preceding claims wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(vii) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing genetic elements, host cells, storage compartments or carriers identified in any of (i) to (vi); and/or

(vii) at least one yeast strain comprising a *can*1 and a *cyh*2 mutation.

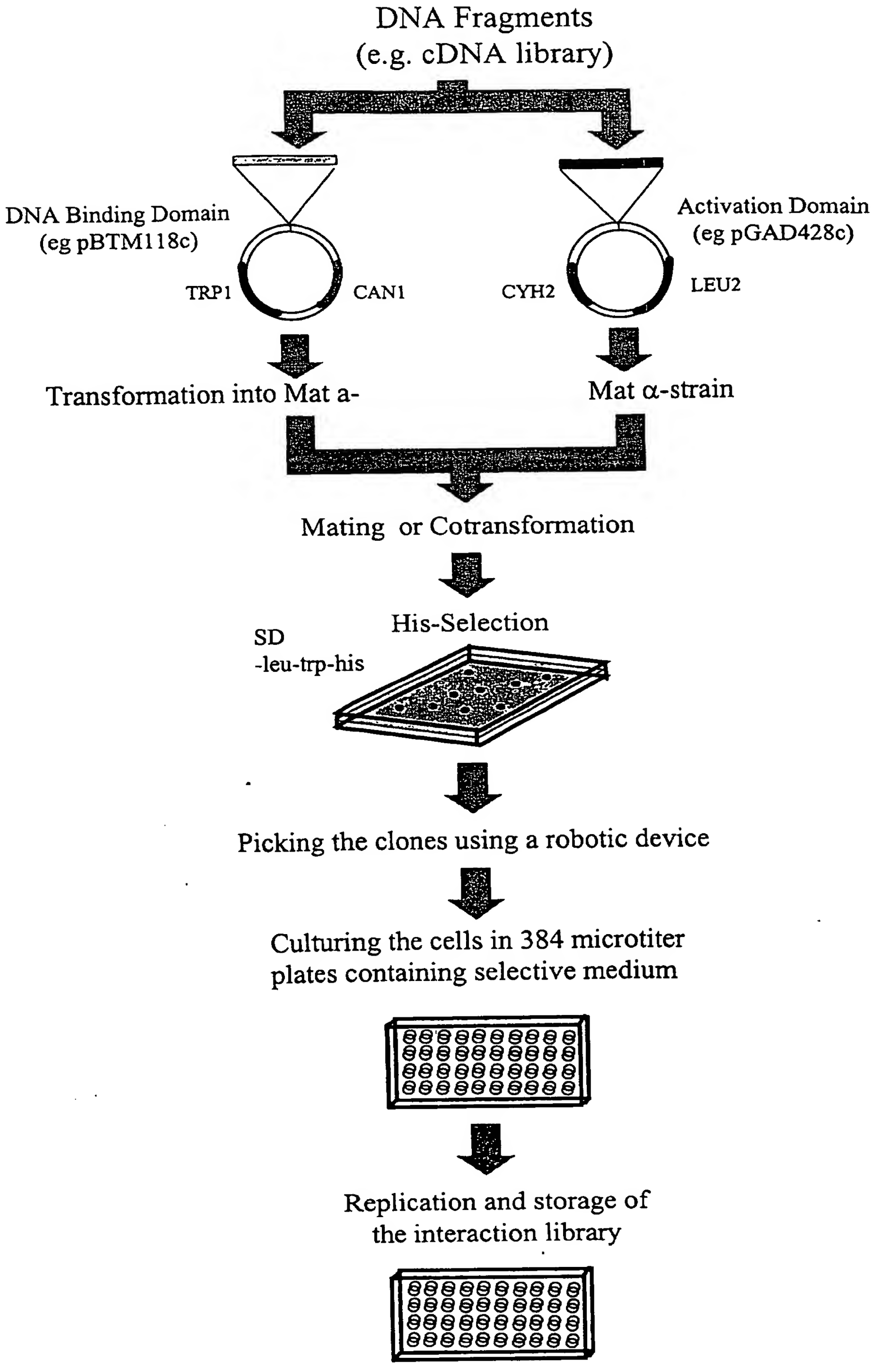
35. The kit of claim 34, wherein said host cells of (i), (ii) or (v) are contained in at least one storage compartment.

36. The kit of claim 34 or 35, wherein said genetic information or said potentially interacting molecules encoded by said genetic information as specified in (i) or (iii) are contained in at least one storage compartment.

ABSTRACT

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to detect positive clones from the rather large numbers of false positive clones isolated by conventional two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules from false positive clones. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules.

Figure 1


DNA Fragments
(e.g. cDNA library)
DNA Binding Domain
(eg pBTM118c)
TRP1
CAN1
Activation Domain
(eg pGAD428c)
CYH2
LEU2
Transformation into Mat a-
Mat α-strain
Mating or Cotransformation
His-Selection
SD
-leu-trp-his
Picking the clones using a robotic device
Culturing the cells in 384 microtiter
plates containing selective medium
Replication and storage of
the interaction library

Spotting of yeast cells onto a nylon membrane;
Transfer of these membranes to different selective media

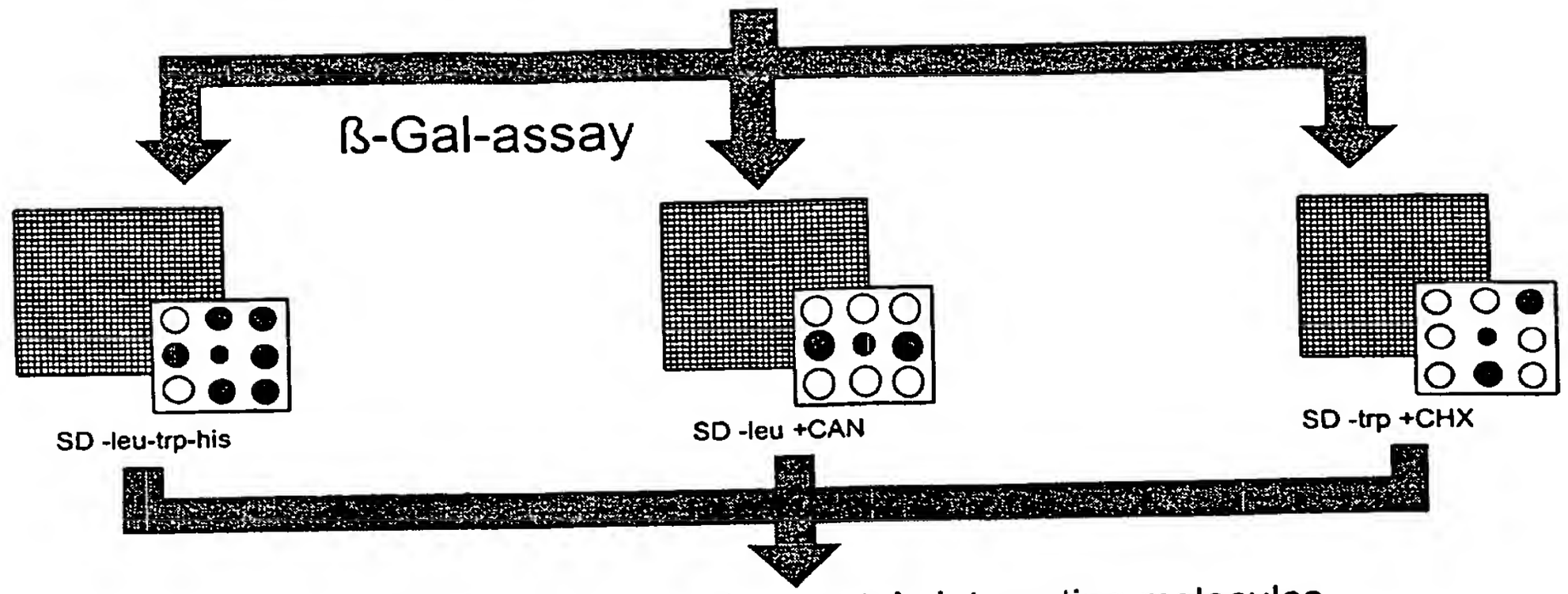


Figure 1 continued

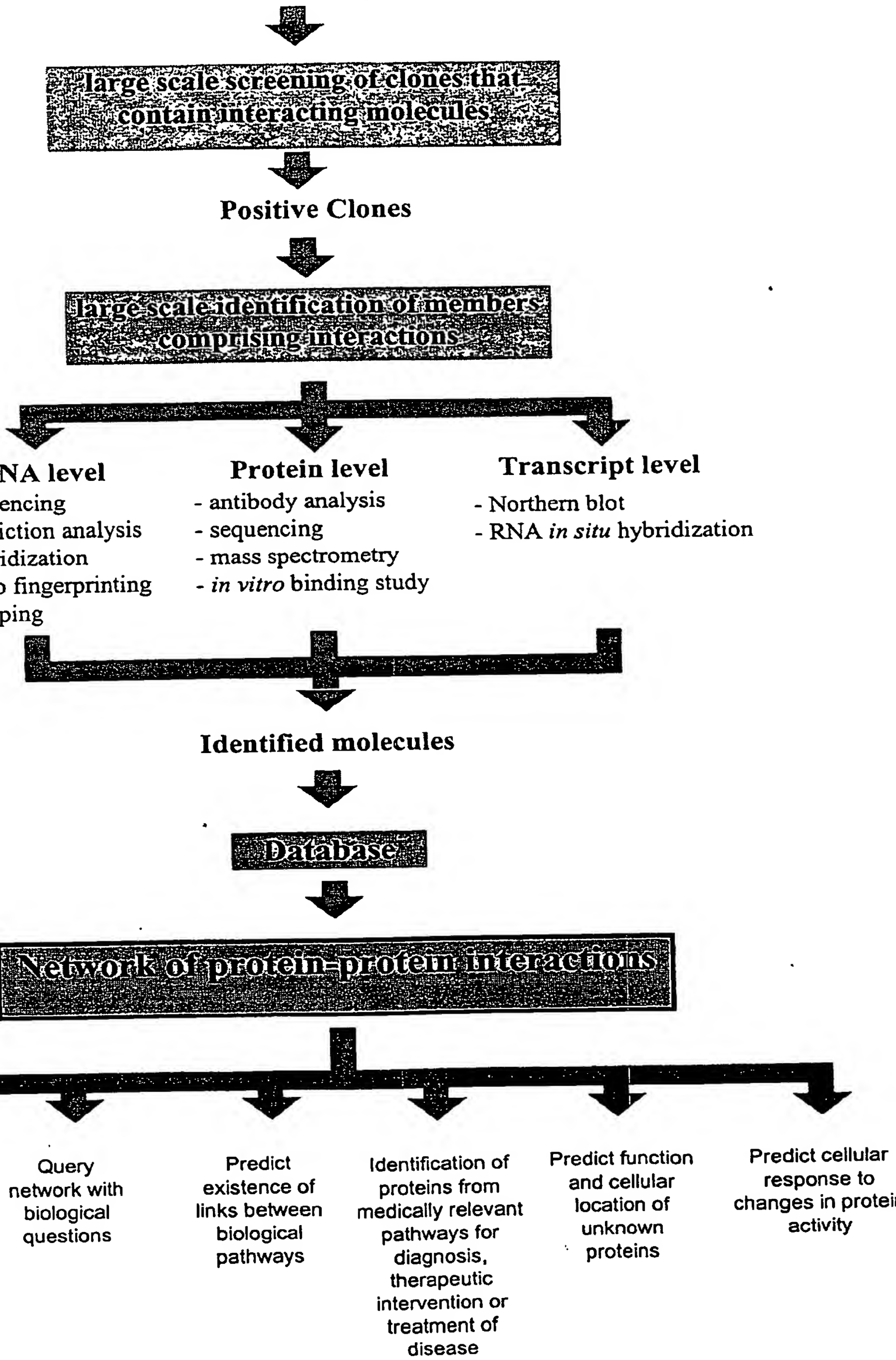
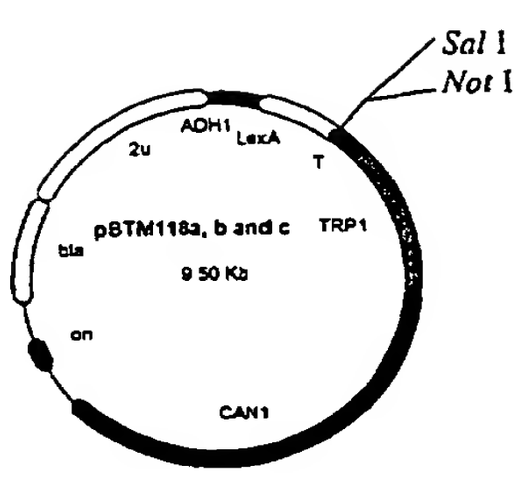
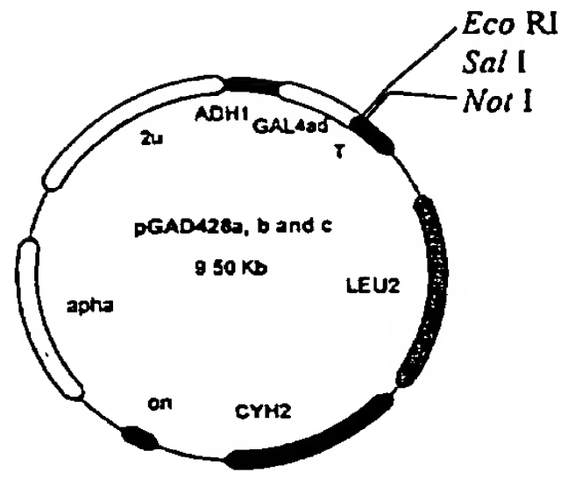

Interaction libraries
large scale screening of clones that contain interacting molecules
Positive Clones
large scale identification of members comprising interactions
DNA level
- sequencing
- restriction analysis
- hybridization
- oligo fingerprinting
- mapping
Protein level
- antibody analysis
- sequencing
- mass spectrometry
- in vitro binding study
Transcript level
- Northern blot
- RNA in situ hybridization
Identified molecules
Database
Network of protein-protein interactions
Develop network with further data
Query network with biological questions
Predict existence of links between biological pathways
Identification of proteins from medically relevant pathways for diagnosis, therapeutic intervention or treatment of disease
Predict function and cellular location of unknown proteins
Predict cellular response to changes in protein activity

Figure 2





a) TCG AGT CGA CGC GGC CGC TAA CCG G
Sal I *Not* I STOP

b) TCG AGG TCG ACG CGG CCG CAG TAA CCG G
Sal I *Not* I STOP

c) TCG AGA GTC GAC GCG GCC GCT TAA CCG G
Sal I *Not* I STOP

Fig. 3a and b

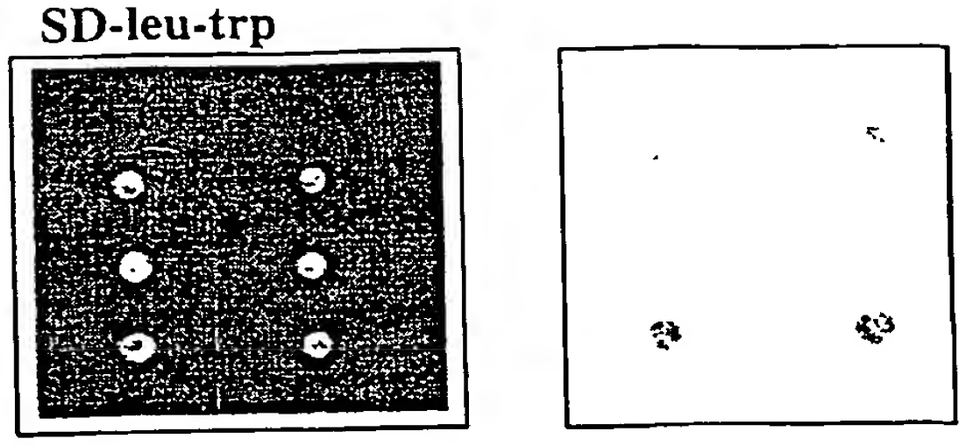
growth ß-gal assay

SD-leu-trp

pBTM117c-SIM1 + pGAD426-ARNT

pBTM117c + pGAD426

pBTM117c-HIP1 + pGAD426

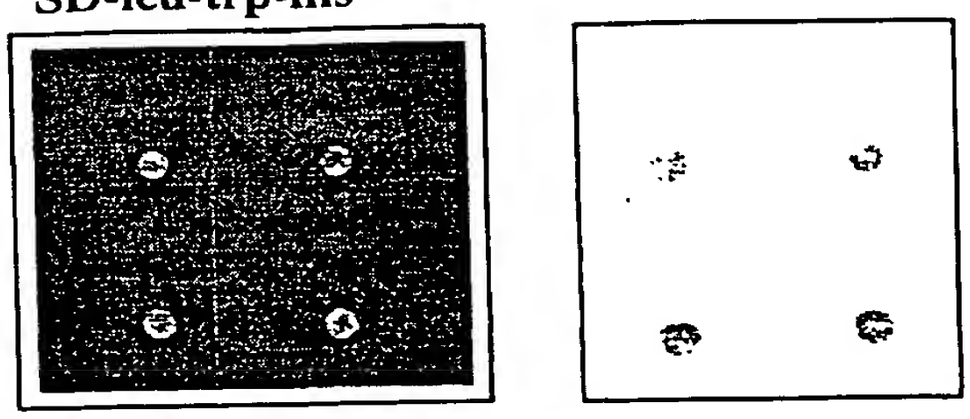


SD-leu-trp-his

pBTM117c-SIM1 + pGAD426-ARNT

pBTM117c + pGAD426

pBTM117c-HIP1 + pGAD426

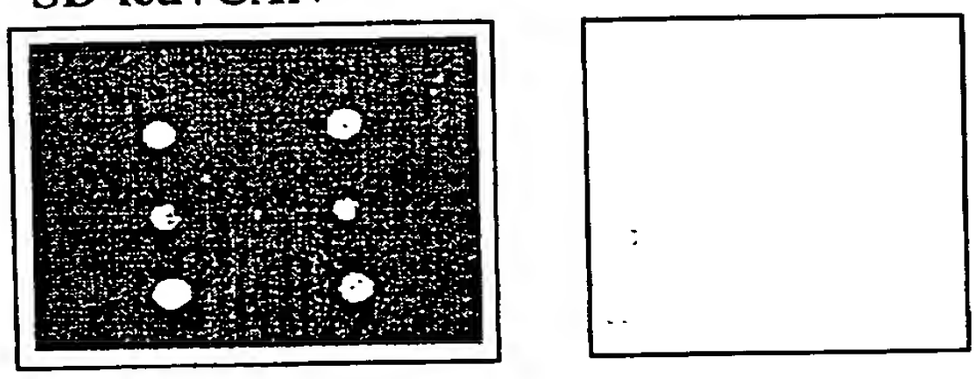


SD-leu+CAN

pBTM117c-SIM1 + pGAD426-ARNT

pBTM117c + pGAD426

pBTM117c-HIP1 + pGAD426



SD-trp+CHX

pBTM117c-SIM1 + pGAD426-ARNT

pBTM117c + pGAD426

pBTM117c-HIP1 + pGAD426

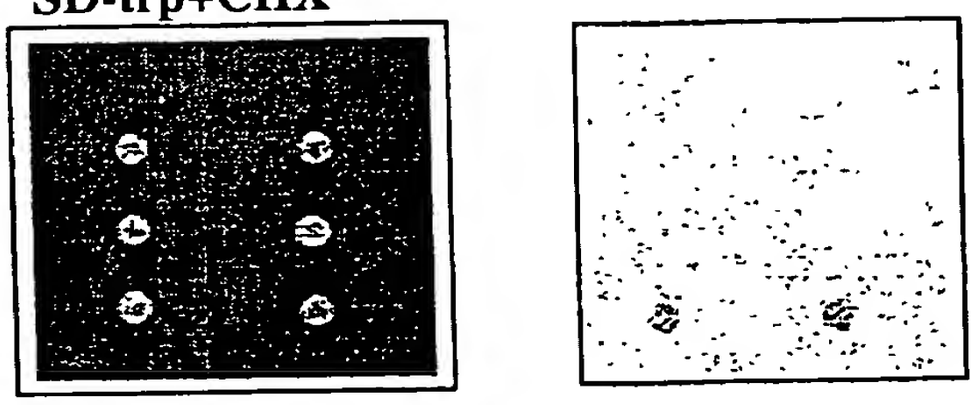


Figure 4

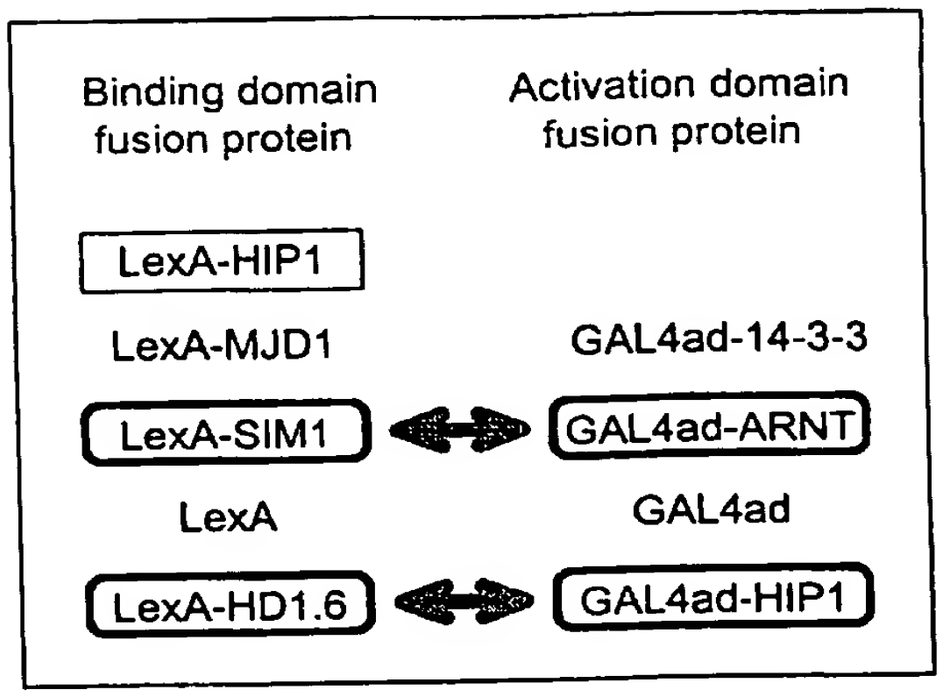
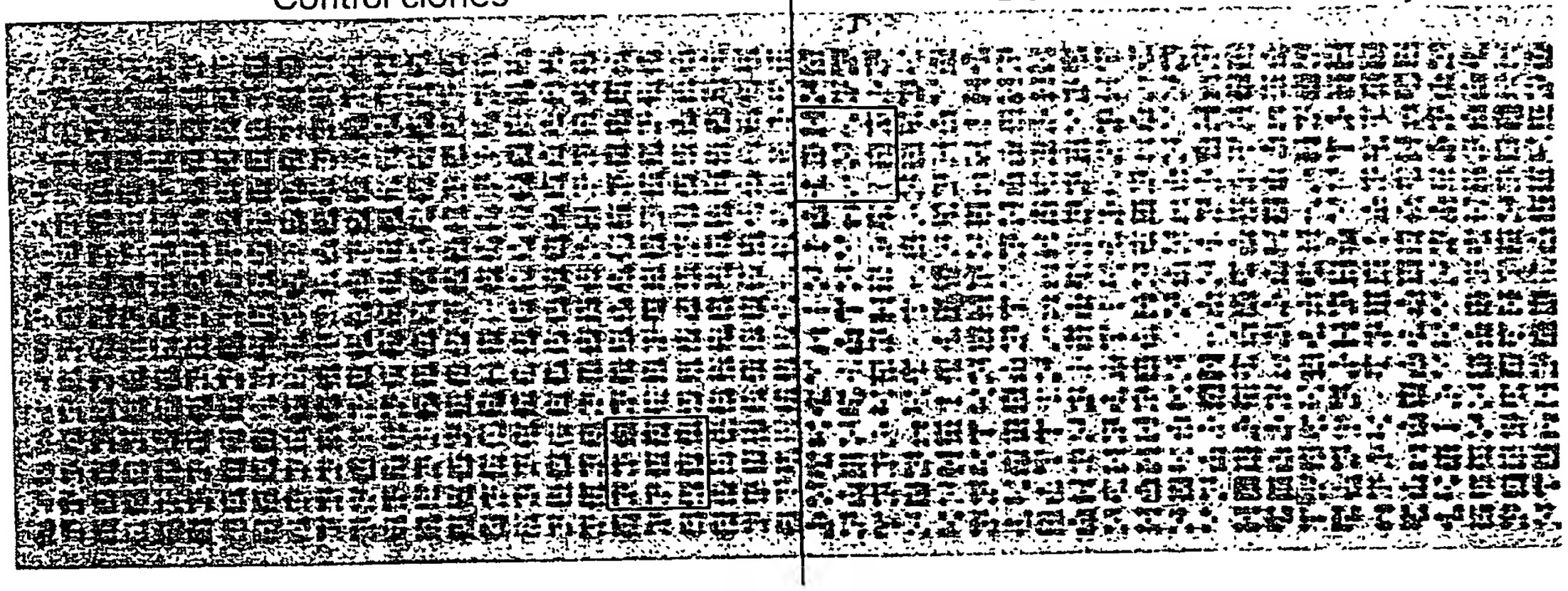

Binding domain fusion protein
Activation domain fusion protein
LexA-HIP1
LexA-MJD1
GAL4ad-14-3-3
LexA-SIM1
GAL4ad-ARNT
LexA
GAL4ad
LexA-HD1.6
GAL4ad-HIP1

Figure 5

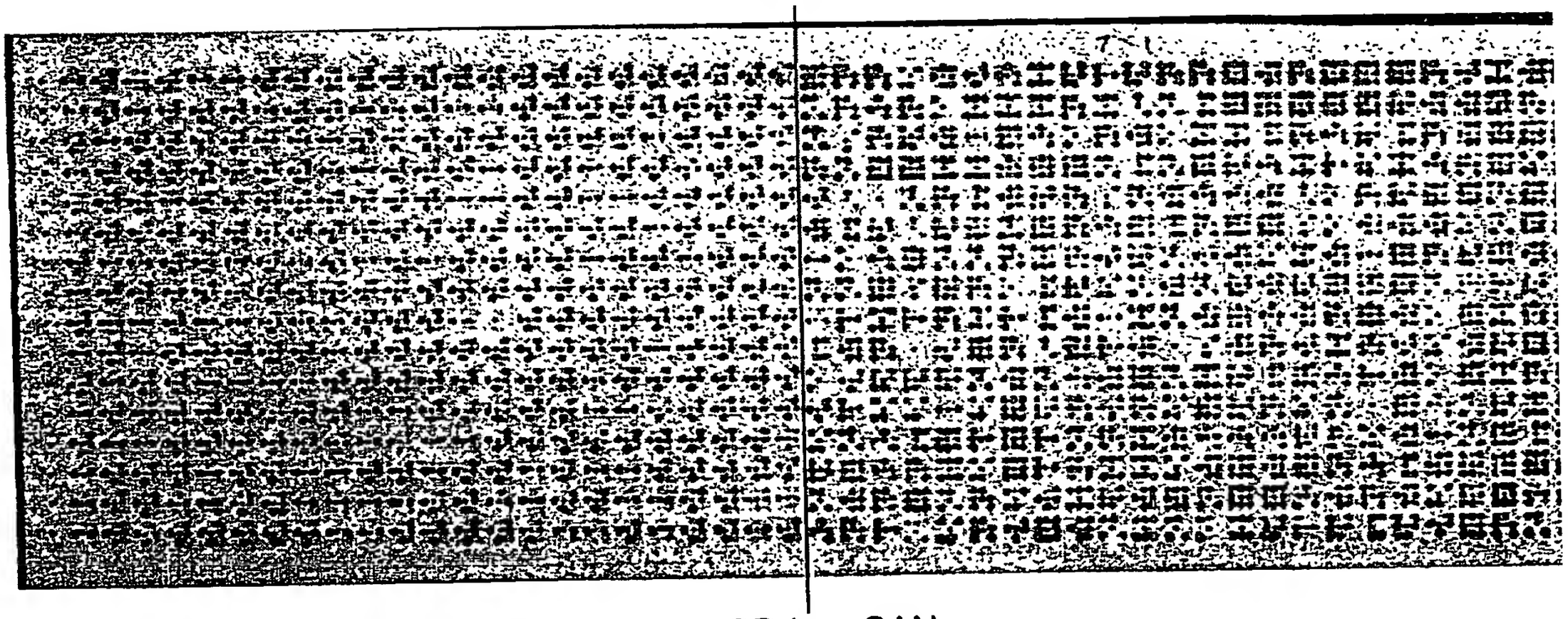
a. SD-leu-trp-his

Control clones

Defined interaction library



b. SD-trp+CHX



c. SD-leu+CAN

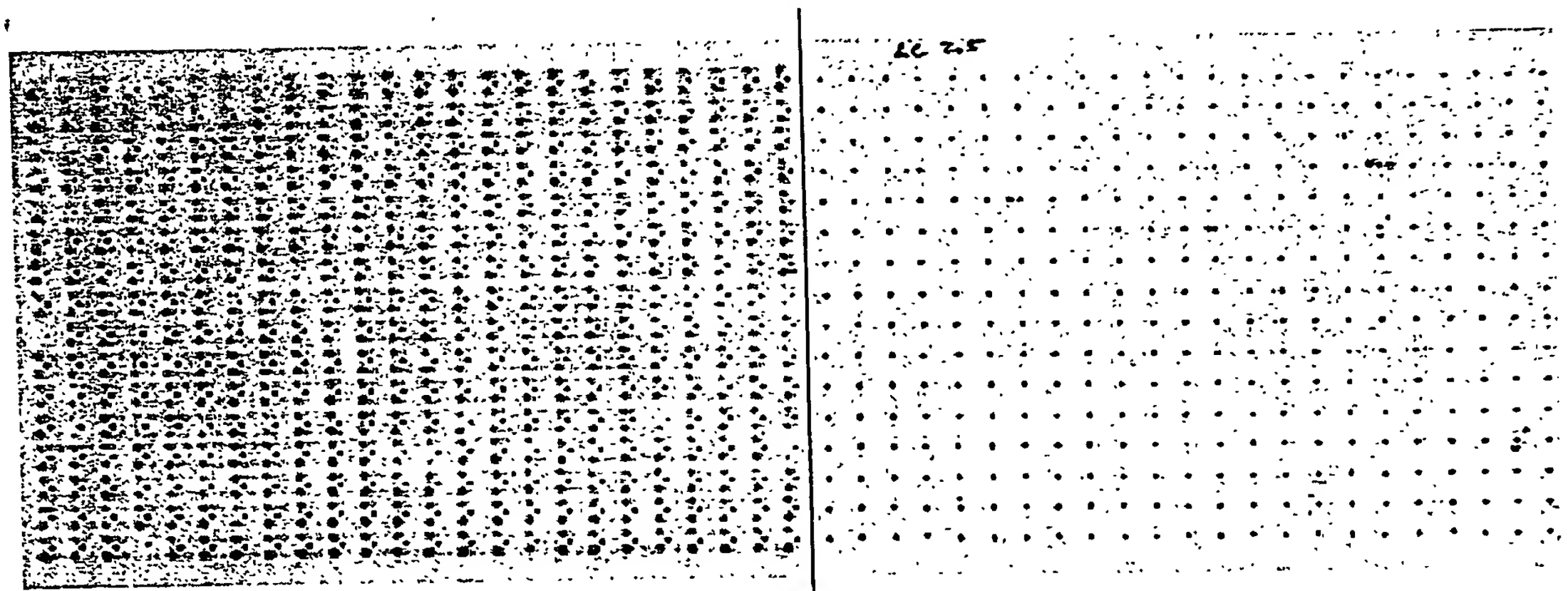
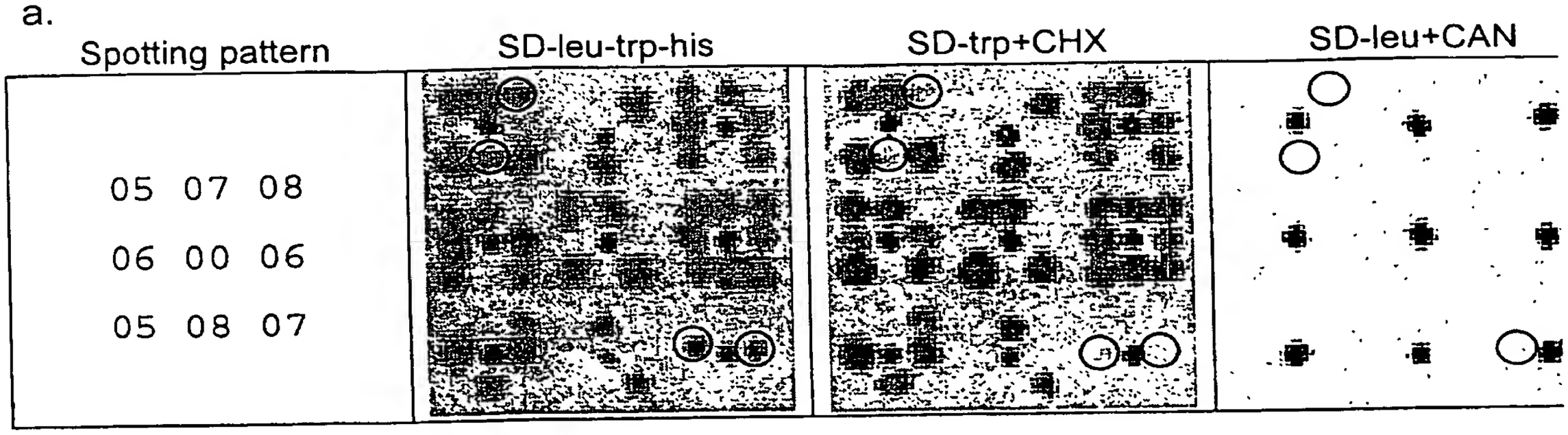


Figure 6

a.


Spotting pattern
SD-leu-trp-his
SD-trp+CHX
SD-leu+CAN
05 07 08
06 00 06
05 08 07

b.

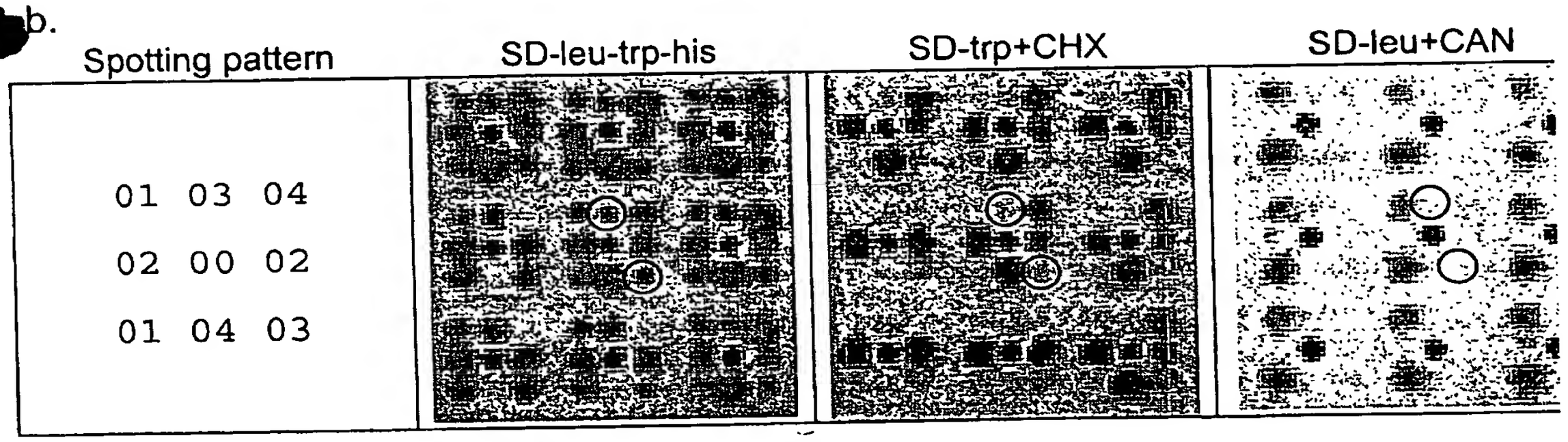

Spotting pattern
SD-leu-trp-his
SD-trp+CHX
SD-leu+CAN
01 03 04
02 00 02
01 04 03

Figure 7

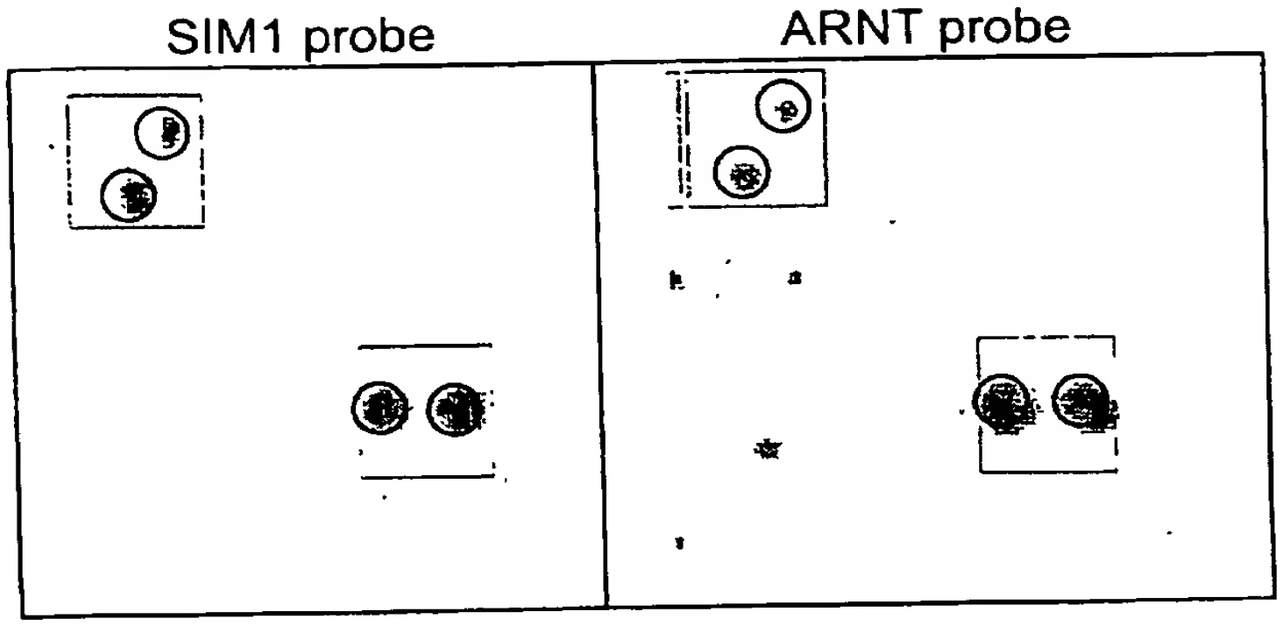


Figure 8

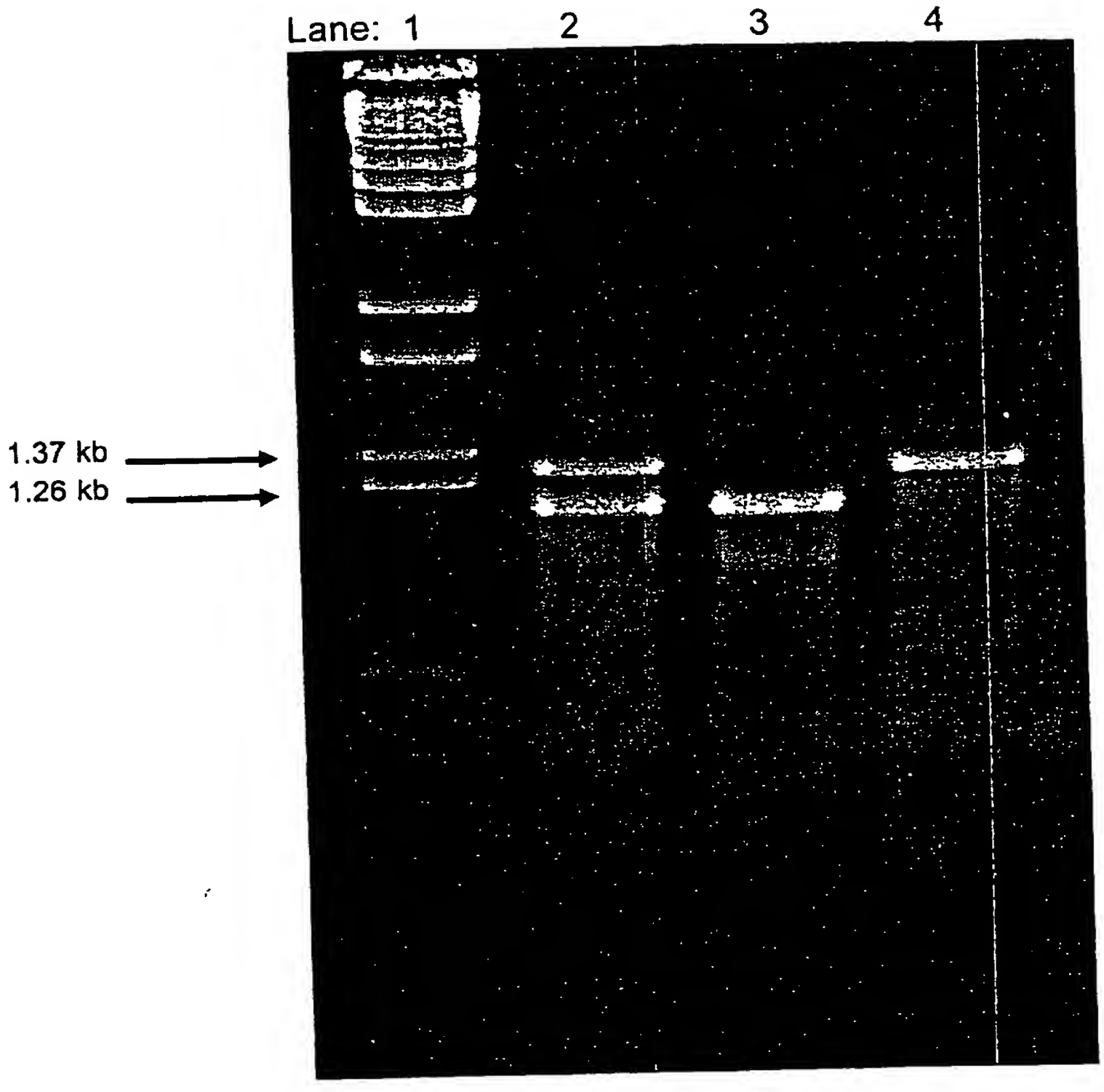


Figure 9